THE UNIVERSITY OF ALBERTA

INFORMATION PROCESSING ON CORRECT RESPONSE TRIALS IN CONCEPT IDENTIFICATION

by

© James Joseph Forest

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

AND RESEARCH IN PARTIAL FULFILLMENT OF THE

REQUIREMENT FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

EDMONTON, ALBERTA

SPRING, 1973

ABSTRACT

One hundred and twenty Ss solved two regular concept identification problems in an experiment designed to test the Bower and Trabasso assumption that Ss cannot logically modify their learning rate on the basis of information available on correct response trials. A novel measure of information processing was used which assigned to each concept a score that represented the degree to which it would correctly categorize a special series of correct response trials. These consistency scores formed a theoretical distribution from which Ss sampled after correctly categorizing a series of stimulus cards. Results indicated that Ss did not sample from this distribution at random. Consistency scores for concepts improved as the number of correct response trials increased. Furthermore, there was an improvement in the average consistency score over the two problems. Consistency scores were related to the degree of difficulty of the logical rule used to categorize the correct response trial stimuli. Although there were certain statistical and methodological limitations on the experimental conclusions, the results of this study suggest that the

Bower and Trabasso conclusion is wrong. Subjects can process some relevant information from a series of correct response trials which will modify later behavior.

ACKNOWLEDGEMENT

I would like to express my sincere appreciation to Dr. W. Runquist for his comments and criticisms which greatly assisted in the development of this thesis and as well, to thank him for the patience he has shown while it was being written.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES CONT.

LIST OF FIGURES

LIST OF FIGURES CONT.

INFORMATION PROCESSING ON CORRECT RESPONSE TRIALS IN CONCEPT IDENTIFICATION

Present day investigators in the area of concept identification have, for the most part, adopted the view that man is an information processing organism. Furthermore many researchers have made the implicit or explicit assumption that humans are active seekers of information, who use a variety of strategies to achieve their goals. While this type of cognitive approach is open to all the weaknesses and criticisms that can be put forward by behaviorism and a deterministic psychology, it has motivated enough empirical and theoretical research to deserve further investigation. This thesis proposes a method of investigating an integral part of the process of concept identification involving a subject's ability to successfully abstract and interpret potentially useful information. Specifically this will involve an attempt to determine if subjects can use the relevant logical information available on correct response trials to significantly alter their concept identification performance in some fashion.

It is quite clear that concept learning systems with any number of built in limitations may be constructed to study various relevant theoretical questions. There do

seem to be, however, certain basic discriminations which must be incorporated in every concept learning system if it is to show even minimal performance in concept attainment. First, it would appear that there must be some type of recognizable goal along with a criterion for judging when that goal has or has not been achieved. In addition there must exist a source of information and some form of feedback related in at least a partially consistent fashion to the various alternative information states. Usually these factors are specified through experimental control and instructions. Thus, the task for the concept learning system becomes primarily one of successful discrimination of the states of the source of information in relation to the available responses as specified by the feedback. For a problem to exist, the system must contain at least two different response producing mechanisms(hypotheses), one of which is capable of satisfying the criterion. Once the system has selected a hypothesis and made a prediction of what the correct response is (whether this involves the categorization of a stimulus or a statement of the potentially correct hypothesis specifically selected), there can only be three possible outcomes. The prediction can be confirmed, infirmed, or receive no feedback whatsoever. If one

considers only the more common types of non-verbal concept attainment problems, it is quite obvious that there are important logical differences in these categories from the experimenter's point of view, as well as for the learning system whose potential information processing strategies are limited by its degree of sophistication. Computer systems with specific limitations have been designed which make optimal use of the information in confirmed and infirmed feedback stimuli at least for simple concept problems. With human systems however, the situation appears to be quite different. Not only are the basic limitations poorly understood in relation to these three types of feedback, but subjects seem to possess a wide variety of strategies varying in efficiency, which they apply depending on their subjective estimates of the task requirements. Since subject performance is so highly dependent on these strategies in terms of the amount, type and accuracy of information they provide for future hypothesis testing, it seems reasonable to attempt to obtain a better understanding of how subjects approach and deal with feedback of various kinds. The following discussion will attempt therefore to present some of the unanswered and controversial questions existing in this area as well as provide a partial review and evaluation of the empirical data known to date about these types of feedback.

Blank Trials

From a logical point of view, the type of responses least useful to subjects are those which receive no feedback and thus supposedly prevent the occurrence of information processing on that trial. Presumably the only thing that the subject could learn under these circumstances was that the stimulus presented was a legitimate part of the stimulus universe.

One of the first theorists to incorporate the idea that blank trials produce no significant behavior changes was Restle (1955). His mathematical theory of discrimination learning hypothesized two types of cues: relevant cues (r) which were correlated with the correct categorization of stimuli and irrelevant cues (i) which were randomly associated with correct categorization. According to Restle, it was only on reinforced trials that subjects could learn whether cues were relevant or not. Despite contradictory statements in his discussion, Restle did not suggest that irrelevant cues were eliminated from consideration through consistent non-reinforcement (non-reinforced trials) but rather that they were eliminated through inconsistent association with positive or negative reinforcement.

The follow up study by Bourne and Restle (1959) stated the effects of non-reinforced trials more clearly by showing that when the parameter ($\theta = \underline{r} / (\underline{r} + \underline{i})$) measuring learning rate is equal to zero the expression $(1 - \theta) + \theta$ drops out of the conditioning and adaptation equations. These were the equations used to predict the theoretical rate at which a relevant cue would become conditioned to the correct response or an irrelevant cue would become neutralized or adapted. In terms used by Bourne and Restle this meant that the probability that cue $\underline{k}$ was conditioned or adapted on trial N + 1 when no reinforcement is given, was equal to the probability of that cue having been conditioned or adapted on trial N (i.e. no learning could have occurred on trial N + 1). Levine (1963, 1964) expanded on this idea of blank trials and developed a technique for discovering the hypothesis the subject was tracking on any particular trial in a concept attainment problem. This was done by placing a specially constructed series of blank trials after each feedback trial so that no matter what stimulus attributes were selected from the feedback trial for testing, they would produce only one positive classification in all the blank trials. By checking which blank trials were classified as positive, Levine could determine the hypothesis

being used. The results of his experiments indicated that the insertion of a set of blank trials (non-reinforced trials) between any two feedback instances or non-instances had no effect on the categorization of these or other stimuli. This technique of using blank trials has been employed quite extensively by other experimenters in an attempt to avoid many of the problems involved in trying to relate verbalized hypotheses to the actual stimulus categorizations made by subjects.

The presence or absence of statements on the effects of blank trials on learning is more obvious in some of the recently formalized mathematical theories of concept identification. Falmagne (1970) has proposed a model which gives each hypothesis a strength or subjective plausibility, using the assumption that learning is the change in the strengths resulting from information provided by the experimenter. This clearly suggests no changes in hypothesis strength (learning) if no information (presumably feedback) is available. While there is no explicit statement to this effect in the model's strength axioms, the fact that Falmagne allows for the change in the strength of a hypothesis on trial N only when that hypothesis has been confirmed or infirmed on trial N-1 seems virtually to mean the same thing. Nahinsky (1970), in a mathematical theory

attempting to predict stimulus classification in conjunctive problems, specifies seven sampling and replacement axioms none of which indicates the effects of blank trials on performance. However in his discussion of parameter estimation he does quote the Levine, Leitenberg and Richter study (1964) to indicate that the subject's hypotheses about the solution remain quite stable during blank trials and hence determination of sampled hypotheses by this technique should be fairly reliable. Nahinsky cautions against using lengthy series of blank trials however, because of possible distortions over time.

In summarizing the results from these more recent theories and experiments as well as the older ones, two tentative conclusions could reasonably be suggested. The first would imply that subjects can neither move into the learned state (the all-or-none theorists use the term absorption state) nor condition or adapt cues (conditioning theorists) on blank trials. The second conclusion, somewhat more general than the first, would indicate that on a series of blank trials no significant behavior modifications could occur to alter a subject's normal hypothesis testing behavior. While such conclusions would definitely simplify theorizing about concept attainment if completely true, it now appears that certain modifications are required in

them and that the rather innocuous nature assigned to blank trials is perhaps undeserved.

As early as 1958, Bourne and Pendleton (1958) found that groups of subjects who had blank trials inserted in their concept attainment problems learned more slowly, implying that blank trials have a disruptive effect on the learning trials which were reinforced. Bourne and Restle (1959) suggested that this was possibly caused by a stimulus trace which dissipates exponentially during the interval between the end of the pattern presentation and the onset of reinforcement. Since learning can only take place at the time of reinforcement, a variable number of cues would be conditioned depending on the length of delay for reinforcement. The result would be that reinforcement would affect not only the stimulus presentation for which it was intended, but also any cues from previous blank trials which had not decayed. This effect would be greatest for the cues from the immediately preceeding blank trial and maximally disruptive for the case where the blank trial and the feedback trial contained no stimulus attributes in common. An even more interesting extension of this line of reasoning would make the assumption that the reinforcement trace dissipates exponentially over time as well. The result would be that reinforcement

for a feedback instance would persist for a time and possibly overlap a following blank trial. Many of the stimulus cues of the blank trial would then be conditioned to the reinforcement given on the preceeding feedback trial. Such a process would account nicely for Nahinsky's findings (Nahinsky and McGlynn 1968; Nahinsky, Penrod and Slaymaker 1970; Nahinsky and Slaymaker 1969) that the first two blank trials in a series of three following a positive feedback instance tend to be classified positive more often than the third. If, as Nahinsky states, the blank trials formed an orthogonal series with each blank trial testing one of the three possible hypotheses available on the feedback instance, then each subject should classify only one blank trial positive and the distribution of positive classifications should be spread evenly over the three instances. However, data on 90 subjects from one experiment (Nahinsky and Slaymaker, 1969) showed a frequency distribution of 84, 69 and 25 positive responses to the 1st, 2nd and 3rd blank trials respectively. Such a distribution must be considered unusual even if one allows for a possible saliency effect. Nahinsky's explanation for these results is that subjects are sampling more than one hypothesis at a time for testing. This is a reasonable possibility and one which will be considered in somewhat more detail later on in the discussion.

The question of multiple hypothesis testing has been studied by other investigators including Andrews, Levinthal and Fishbein (1969) who approached the blank trial problem from a slightly different angle. While agreeing with Levine (1966) that there is probably no memory loss for the active hypothesis during the blank trial period, they believed that subjects may suffer some loss for what they call secondary hypotheses. The primary hypothesis was defined as the hypothesis tested on the blank trial sequence while secondary hypotheses were defined as those values of the stimulus consistent with the primary hypothesis on the stimulus trial immediately following the blank trial set (i.e. the stimulus values shown to be wrong along with the primary stimulus values when the experimenter says wrong to the subject's classification on the feedback trial). In fact, the experimental data supported this contention. There was a greater tendency for subjects to retest disconfirmed secondary hypotheses as compared to primary hypotheses. One explanation the authors offer, is that possibly the blank trial technique required the rehearsal of the primary hypothesis for all blank trials and feedback trials while a secondary hypothesis is only really rehearsed on feedback trials thus giving differential strength for recall.

This problem is much less important however than the fact that serious questions are now being raised about Levine's work (1963, 1964, 1966) and his conclusion that on blank trials subjects show very few response sequences which are not consistent with some permissable hypothesis. Chumbley's data (1969) for example, indicated that in some circumstances subjects could emit sufficiently large numbers of uninterpretable response sequences so as to make data analysis and interpretation difficult. Nahinsky (1968, 1969, 1970) no longer even considers the question of consistency of response sequences when monitoring conjunctive concepts on blank trials. Almost all subjects show a tendency to classify at least two stimuli as positive even though they overlap on only one attribute. While these studies do not disprove Levine's conclusions, they do raise some important questions not only about blank trial usage, but also concerning the strong support Levine's studies supposedly provide for the belief that man is a hypothesis tester. It just may be that the blank trial approach is not as easy or consistent a method for following subjects' hypotheses as originally believed.

One other study deserves mention in this discussion of blank trials if only because the results obtained

were directly opposite to those found by Bourne and Pendleton (1958). In a study of knowledge of results, Cahoon (1970) varied the frequency of feedback and found that card choices to solution decreased as the frequency of feedback decreased. He suggested that subjects with lower frequencies of feedback made better use of the information available to them at least after the point at which they had enough information to solve the problem. It would appear unreasonable, without further evidence, to assume that there was some methodological problem involved in the experiment. Yet, there is no other obvious explanation for the improved performance observed in this study as opposed to the decrements found elsewhere or the no change results obtained in most experiments.

Any theorist attempting to draw these seemingly divergent results and conclusions into a coherent model is undoubtedly going to be faced with the problem of too many unanswered questions and too few definitive statements. This author favors the Levine conclusions because of the extensive and basic background research (1963, 1964) carried out prior to the publication of the blank trial technique. This does not deny the likely possibility that changes will be required as the number of studies

using this approach increase and possibly validate the research quoted here. In any case, it is unlikely that subjects will react correctly to blank trials until such time as instructions are devised which can adequately stress the lack of logical information on these trials. The only way to improve on these tentative statements will be through additional direct research on blank trials although it does not seem likely that it will be forthcoming soon. The majority of work is being concentrated on what experimenters generally feel are the more basic and informationally important categories of confirmed and infirmed predictions.

Error Trials

For a variety of theoretical and empirical reasons, confirmed and infirmed predictions have not generated equal interest or research. Confirmed predictions were relegated to a secondary role until recently for two reasons. First, most investigators seemed to feel that this type of prediction required a very simple strategy for maximum performance: namely, when a cue or hypothesis predicts correctly, continue to use it. Such a simple rule made it extremely difficult to detect any changes caused by confirmed predictions, especially if summary statistics such as trials to criterion or errors were used. Furthermore, any attempt to explain how such predictions about concept identification were related to summary statistics required a long line of intervening variables or hypothetical constructs. Infirmed predictions on the other hand, seemed to offer a much easier approach with a greater change of explaining the major factors affecting concept learning performance. A variety of strategies could be adopted on error trials and they differed considerably in terms of the amount and kind of information they provided (Gregg and Simon, 1967). Thus it seemed relatively easy to establish empirical relationships between the dependent variables and various strategies as well as detect changes when errors were systematically

varied. Also the changes were clearly behavioral and did not require the postulation of unknown mechanisms, a position favored by behaviorists who studied concept attainment.

Actually the first few theoretical papers on concept attainment (Restle, 1955; Bourne and Restle, 1959) emphasized the importance of both confirmed and infirmed predictions by their assumption that learning occurred on all trials through conditioning and adaption of relevant and irrelevant cues. This approach was dropped rather quickly by Restle, who, in an attempt to correct certain errors in his former method, ended up with a totally new idea. This theory (Restle, 1961, 1962) concentrated on strategies (consistent patterns of responses), their selection and how they were related to errors. There were two important assumptions (Restle, 1961): (1) when a response was correct, a subject continued with the same strategy, (2) if the response was wrong, a subject re-sampled with replacement from the set of available strategies. This meant that errors were uncertain re-current events because they always set the subject back to zero in the sense that all events preceeding this error were statistically independent of all subsequent events. Furthermore, if an error had occurred at trial T, then

the conditional probability of it occurring again on trial T + N was equal to the probability that an error occurred on the Nth trial. The model specifies that errors depend on the proportion of wrong, irrelevant and correct strategies in the total pool. Later these assumptions were used to develop a somewhat new stochastic model which took into consideration the number of strategies a subject might be testing when he resampled after errors (Restle, 1962). Subjects could, for example, sample one hypothesis, M hypotheses (where $M < N$ and N is equal to the total number of hypotheses) or N hypotheses. Restle showed that while these models are not generally alike in the probabilities they use, they would give indistinguishable sequences of errors. These models all propose that subjects resample a certain number of hypotheses on an error and that nothing is carried over from correct or error trials to modify the probability of sampling any particular hypothesis on subsequent errors. This meant the the probability of success on trial N + 1 following an error or success on trial N did not change until the criterion run.

Restle's 1962 model of concept attainment did not stimulate much addition research. It did, however serve as a basis from which a number of other models were

developed. The most successful theory to arise from Restle's work was one constructed by Bower and Trabasso (1963 b). It received a detailed theoretical and empirical examination over a period of years, and as a result, underwent several major revisions (Trabasso and Bower, 1964, 1966) before it reached its current form (Trabasso and Bower 1968). Because the theory has proven highly testable and directly relevant to an understanding of correct response trials, it has been examined in considerable detail in the following discussion. The purpose of this examination is not the total rejection of the Trabasso and Bower model but rather the re-evaluation of two of their assumptions on the basis of new theoretical and empirical data. This evaluation is applicable to any other theory making the same basic assumptions.

There appear to be two basic axioms central to the Bower and Trabasso model. Axiom One states that on each trial a subject can be in only one of two states, _L_ or _UL_. A subject in the _UL_ state does not know the correct concept and therefore makes an incorrect response with the probability _p_. When a subject is in state _L_ however, he knows the correct concept and always makes a correct response. The second axiom indicates that after each correct response the subject remains in his previous

state, while after an error he shifts from UL to L with probability r. The combination of these two statements results in errors being defined in virtually the same way as Restle defined them. On the surface Bower and Trabasso appeared to have strong evidence supporting their theory and definition of errors. In one study (Bower and Trabasso, 1963b) they manipulated the chance probabilities of an error by varying the number of response possibilities. If learning occurred only on error trials then a group required to classify stimuli into one of four categories would make errors more frequently than a two category group and hence reach the trial of last error sooner. This in fact occurred even though total numbers of errors were equal and consequently learning rates for the two groups did not differ.

Further support for the belief that learning occurred only on error trials came from a number of experiments on reversal learning which also manipulated error position. The basic idea here was to test the same assumption: that after an error a subject was in the same state as he was at the beginning of the problem and that the intervening trials had no effect on the probability of solution. The general results of these studies (Bower and Trabasso, 1963a, 1963b; Trabasso and Bower, 1964a)

showed that when groups undergo reversals of S-R assignments after every other error, they still make the same number of informed errors before learning as controls trained with fixed S-R assignments. Even a shift between dimensions produced roughly the same number of errors (slightly less, because on a shift the experiment may switch the correct response assignments into line with the hypothesis the subject is currently using). Bower and Trabasso concluded from these studies that the effective information promoting learning, occurs only on informed errors and that the probability that a subject solves per error appears to be unaffected by the past history of inconsistent reinformcement to the cue on which he solves the problem.

In a somewhat different examination of the presolution stage of concept identification, Bower and Trabasso hypothesized that according to their theory, there should be no evidence of learning prior to the final error (stationarity) and that trial to trial responses should be independent (independence) if errors are uncertain events, and no learning occurs on correct response trials. These predictions received wide support under a variety of experimental situations (Bower and Trabasso, 1963a; Erickson and Zajowski, 1967; Erickson, Zajowski and Ehmann, 1966; Trabasso and

Bower, 1964a; Trabasso and Bower, 1966). In addition, it was shown that many important variables, including distribution of total errors and trial number of last error could be predicted with reasonable accuracy from a simple mathematical model drawn from their assumption.

This evidence seems to provide strong support for the conclusion that error trials set subjects back to zero and that learning occured only following error trials. Again however as in the case of blank trials, evidence was accumulating to indicate that such conclusions might be wrong, particularly in regard to the first one which turned out to be more easy to test empirically.

Levine (1962) carried out a two phase experiment using random reinforcement presumably setting the subject back to zero. After this, the subjects shifted without any noticeable break into phase II which was a regular concept identification problem having one consistently reinforced answer. Levine found that the group receiving zero random reinforcements (RR) had the fastest learning rate, while groups receiving four or more RR's showed significant decrements in performance. The last error in phase I had not set the subjects back to zero and Levine concluded that possibly the RR caused subjects to ignore or set the relevant cue aside temporarily. Holstein

and Premack (1965) retested this question and supported Levine's findings, however they suggested a different explanation, one consistent with the Bower and Trabasso theory. They suggested that RR caused subjects to form new hypotheses when none of the regular ones worked. Thus the size of the hypothesis pool was increased and the probability of learning or selecting the correct hypothesis after RR was correspondingly reduced. Both the Levine and the Holstein and Premack explanations predicted that the noncontingent wrongs in RR were the cause of the decrements in concept identification performance. Merryman, Kaufman, Brown and Dames (1968) tested this prediction using a procedure similar to that employed by Levine. There were two experimental groups which differed in the reinforcement schedule they received in phase I of the two part experiment. The first group received six "right" feedbacks from E while the second group received six "wrong" feedbacks in phase I. Examination of the results showed that the performance decrements were produced by a change in learning rate not presolution guessing rate or initial performance. Of more importance however was the finding that the greatest decrement occurred for the group receiving six noncontingent "right" reinforcements and not the "wrong" feedback group as predicted.

While the Merryman et al results are interesting,

they were not supported by Erickson (1968) when he used the Levine technique of blank trials for determining a subject's hypotheses. Subjects in the control group solved a concept identification problem in the regular fashion, while subjects in the experimental group were given a wrong reinforcement on the first feedback trial indicating that their initial hypothesis was incorrect. By presenting a blank trial series prior to the feedback trial, Erickson was able to determine what the initially sampled hypothesis was. This hypothesis was then made the solution to the problem after the inconsistent reinforcement on the first feedback trial. If the error set subjects back to zero then once the first feedback trial was ignored, the experimental and control groups should show equal levels of performance. The data showed however that the experimental group took an average of two additional hypotheses to solve their problem indicating that the error caused them to set their initial hypothesis aside temporarily. In support of this finding, Kenoyer and Phillips (1968) used a two part experiment, with part I employing noncontingent reinforcement while part II constituted a regular type of concept attainment problem. The number of wrong reinforcements in part I varied from 0 to 2, and with each increase in number of errors, subject's performance in part II showed a corresponding decrement.

In an interesting variation, Erickson, Block and Rulon (1970) controlled the relationship between the correct hypothesis and the error position rather than manipulating the number of errors. In experiment II of this paper, Ss received "error" feedback on each of the first three hypotheses they tried (only Ss who used three different hypotheses were tested) and then were shifted to a concept identification problem in which one of these three hypotheses was correct. The lag 1 group, whose solution was the hypothesis they tried on the third error block, showed the poorest performance while lag 2 and lag 3 groups showed progressively faster solution rates. This indicated that errors do not set subjects back to zero, because the hypothesis used on an error trial farther back in the stimulus series has a greater chance of being selected than one tested on a more recent error trial.

Bower and Trabasso eventually revised their theory despite the fact that many of their previous theoretical predictions about errors had been supported. They discovered (Trabasso and Bower, 1966) for example, that they had not used an appropriate control group against which to compare their dimensional shift group. A subject in the dimensional shift group could solve on either one of the dimensions that the S-R assignments were being

shifted between. An appropriate control group would allow subjects to learn on two separate and independent dimensions as well. Previously the dimensional shift group had been compared against a control group in which there was only one correct dimension. The previous inappropriate control group only had one correct dimension. When the dimensional shift group was compared against the new control group, the shift group actually showed a significantly lower learning rate. To explain this result, Trabasso and Bower hypothesized that subjects remember the specific stimulus pattern and correct response from trial N-1 and when an error occurs on trial N, subjects compare the two stimuli, temporarily setting aside any attributes which had inconsistent response assignments.

While this revised sampling rule for error trials fitted nicely with the results from previous experiments on reversal learning, it too soon came under investigation and was found to be inconsistent with some data. Using the techniques of misinformative feedback and noncontingent feedback, Rogers and Haygood (1968) discovered not only that subjects changed their hypotheses on errorless trials but that on errors subjects were as likely to maintain their hypotheses as they were to alter them. The fact that there was no significant relationship between informed

errors and hypothesis changes indicated to the authors at least, that there was no simple relationship between hypothesis revision and error feedback. Errors were not setting subjects back to zero, in fact they were not even causing subjects to resample hypotheses with any high degree of reliability.

Dodd and Bourne (1969) also retested the revised sampling rule for errors by investigating three assumptions the Trabasso and Bower (1966) model makes. According to their theory: (1) A hypothesis will be altered only after it has been infirmed by an error. (2) Any new hypothesis which is selected will be consistent with the S-R assignments from that error. (3) A consistency check will be performed between trial N-1 and N where the error occurred and any hypothesis showing inconsistent S-R assignments will not be selected for testing. In both studies that Bourne and Dodd reported, all three assumptions were violated a large number of times. This seems to indicate that while the Bower and Trabasso model is not completely wrong, it has to change its view of errors and correct responses.

While the Dodd and Bourne experiments were primarily concerned with the revised sampling rule, an experiment by Richter (1969) published at the same time,

attacked a more basic foundation in the Trabasso and Bower model. Its primary purpose was to re-examine the multiple reversal experiment (Bower and Trabasso 1963a; Trabasso and Bower, 1964b; Trabasso and Bower, 1966) and show through analysis of the stimulus sequence that such experiments do not allow a strong test of resampling assumptions if summary statistics are used. Indeed this prediction was supported when Richter found evidence for recall of stimulus information from one or two preceeding trials by analyzing stimulus sequence differences. This occurred even though summary statistics indicated no difference in number of informed errors for a reversal group and a control group. While Richter provides no specific discussion on how his data affect the Trabasso and Bower model, it seems clear that in combination with other studies, the evidence against their interpretation of learning and information processing is strong.

Up to this point the criticisms leveled against the Trabasso and Bower theory have been strictly empirical because this obviously remains the ultimate test of any set of predictions. Numerous other experiments (Bourne and Haygood, 1960; Bourne, 1963; Fishbein, Benton, Osborne and Wise, 1963; Johannsen, 1962; Pishkin, 1960) could be examined, but hopefully the ones discussed here

are sufficient to indicate the present underlying uncertainty concerning the nature of information processing following an error. At this point, therefore, a number of theoretical criticisms will be set forth, not with the intention of disproving the theory but rather to show some additional problem areas in it. The implication of this discussion will be that the theoretical status of both errors and correct response trials needs to be revised in terms of their relative importance.

The basic problem with the Trabasso and Bower model seems to have arisen in trying to determine exactly what is meant when it refers to learning, and the specific conditions under which it is believed to occur. The authors state that learning occurs only when a subject goes from the UL to the L state. A subject is assumed to be in state L after a criterion of J correct responses (an optimal J value is not known however). If J is of sufficient length (10 to 30 responses is the general length), it is further assumed that the subject was using one of the correct hypotheses when he achieved criterion and not an irrelevant one. Since at least one error usually precedes the criterion it is also assumed that the error occurred because of an irrelevant or wrong strategy and not by using the correct one and

accidently giving the wrong response. Under these circumstances learning could be defined as any one or a combination of three factors; adopting the correct hypothesis, entering state L or completing the criterion, and at various times Trabasso and Bower seem to alternate between different definitions. This can result in problems; for example, one might conclude that the move from the UL to the L state is learning and that this involves more than just the selection of the correct concept. Conceivably an error could occur, the subject could select the correct hypothesis but with some probability 1-r he does not learn. If this subject is using the correct concept in an appropriate manner he must reach the criterion in the UL state or have learned on a correct response trial. Considering the opposite point of view, if learning is just the selection of the correct hypothesis then feedback is irrelevant and the selection of the correct hypothesis prior to seeing any stimulus cards would involve as much learning as a selection made after the presentation of a stimulus series.

Restle (1962) pointed out an additional complication in regard to learning on error trials. While the subject can potentially select from 1 to N hypotheses (where N equals the total number of hypotheses available)

for testing, it would appear that the Bower ad Trabasso model can only be applied to the case where the subject samples one hypothesis at a time as shown by the following argument. When learning is defined as the selection and retention of the correct concept, then any subject who samples the correct one along with several irrelevant hypotheses should be considered to have learned. However, the subject has no logical way of distinguishing between the relevant and irrelevant ones until he receives additional information, so in one sense he should still be included in the unlearned category. If learning is instead defined as the move into the L state then possibly the subject could go into this state not only with the correct hypothesis but also with several irrelevant ones. Neither of these cases seem to involve what Trabasso and Bower had originally intended their definition of learning to imply. There is a real problem here because if the subjects can work with more than one hypothesis at a time and errors still set subjects back to zero (or even cause him to make a consistency check in which case on the average one half of the dimensions will have inconsistent S-R assignments) it is unclear how he learns in the intertrial interval between error feedback and the next stimulus when he had no information to work with. Correct response trials or the criterion run might seem

reasonable but these are ruled out because the model specifies that learning can not occur on such trials.

This discussion has been aimed thus far at showing the vagueness and inconsistency in the Trabasso and Bower definition of learning. Of equal importance however is the likely possibility that in most respects their definition is too narrow. They emphasize the learning of the correct concept, which is not unreasonable, but in some respects they ignore any ability subjects may have to learn which concepts are wrong. Accepting the idea that learning is only the selection of the correct concept and/or the move to the L state ignores the fact that the selection of any hypothesis (including the correct one) on an error is hardly at random and depends on a subject having successfully recognized and differentiated a variety of factors. This would involve such things as discrimination among the dimensions and their relevance, the values, the responses, the paired association between the two, the manner for rejecting incorrect strategies and so on. In fact learning may not lie in the actual selection of the correct hypothesis but rather in the changes in subjective probabilities associated with smaller factors that in combination direct the selection of specific hypotheses or emission of responses.

Even if the specific theoretical criticisms mentioned above could be overlooked for the moment, there is reason to believe that Trabasso and Bower never could have detected evidence contradictory to the assumption of errors as uncertain recurrent events and as the only trial on which learning can occur. Their experiments were set up such that there were only three general types of hypotheses available: wrong hypotheses which never gave a correct answer or response, correct hypotheses which always did and irrelevant ones which were associated with the correct response at chance level, generally .5. If a subject had learned something on a correct response trial this never showed up as a gradual rise in accuracy of performance because the subject could only switch from .5 to 1.00 probability of correct responding. Thus some information (such as memory for one rejected hypothesis) might increase the chance of a correct guess on an error trial from $1/N$ to $1/(N-1)$ and yet if the subject guessed on an error and got the correct concept, his performance went from 50% to 100% correct immediately. On the other hand if the subject had this same piece of information but failed to guess the correct concept his performance remained at chance level even though theoretically at least he had learned something on correct response trials. This would indicate that in the context of their experimental

conditions, the dependent measures used by Bower and Trabasso could not accurately represent variations in amount of learning on correct response trials even if it existed. It is therefore questionable to interpret stationarity and independence as conclusive evidence that no learning has occurred. It may be that all they have shown is that the response sequences they used were of such a nature that no hypothesis could produce performance at levels other than 1.00, 0.50 or 0.00. The picutre is further confused by the possibility that not only the stimulus sequence but the length of the criterion run as well may affect stationarity and independence. If the criterion length is X, the experimenter must use stimulus sequences for which no irrelevant hypothesis can produce more than X-1 correct responses. By suitable selection of stimulus cards, E can vary this down to X-K (where K goes from 1 to X-1) and potentially create or destroy stationarity and independence without being aware of it.

While obviously none of these arguments about stationarity, independence, etc. is conclusive, they do lead one to belive that the confidence Bower and Trabasso placed in their conclusions about error trials and correct response trials was unwarranted. Certainly if the study

of correct response trials is to be any more than a minor academic question, it has to be shown that they have significant efeect on final concept identification performance. First, however, an important obstacle had to be overcome. It was necessary to prove that the partially supported premise of errors setting subjects back to zero and negating the effects of all previous errors and correct responses was wrong. The preceding discussion on infirmed predictions was therefore an attempt to provide sufficient theoretical and empirical arguments to show the weaknesses of the premise that errors are uncertain recurrent events. Once this was accomplished it seemed reasonable to reopen the study of correct response trials and the possibility that subjects may process information on them. The importance Bower and Trabasso have placed on error trials may not be out of proportion, in fact it may be that such trials are more important than correct response trials in explaining the variance observed in concept identification learning. However the evidence which has been accumulating does not support the contention that each new error eliminates the effects of all previous trials. In the third section of this paper, the discussion will turn to a concentration on research more directly concerned with showing the effects of correct response trials. Several of the papers in this

area appear to suffer from various theoretical and methodological problems, but since the amount of research here is so meager they will be considered even though it has to be kept in mind that their results are only suggestive.

Correct Response Trials

The current status of correct response trials as sources of relevant information for concept attainment is rather ambiguous. Recent theoretical articles (Falmagne, 1970; Nahinsky, 1969; Williams, 1971) would seem to suggest that information processing on correct response trials is an experimentally established fact. Closer examination of these studies reveals, however, that such is not the case. Falmagne (1970) introduced correct response trials as important, proposed to study them, and then, in his methods section completely ignored any mention of them. William's study (1971) on the other hand failed to provide any type of dependent measure that would respond specificly to variations in information processing on correct response trials. Consequently it is impossible to separate the effects of correct response trials from errors. Lastly, Nahinsky's (1969) study, involved experimental problems which were: too easy, had possible saliency effects and used blank trials.

This rather strange state of knowledge regarding the nature of correct response trials appears to have arisen because the papers dealing with this type of trial contained undetected methodological faults. The first and perhaps most extensive piece of research carried out on

confirmed predictions involved a set of three experiments (Richter, 1965) which unfortunately were not published. Using a technique whereby different numbers of irrelevant dimensions separate off from the relevant dimension in a set of cards with internally ordered orthogonal dimensions, the author attempted to investigate the role of outcomes ("right" and "wrong") as well as the effects of stimulus sequences on concept identification learning. Each of 25 problems was set up such that four successive trials were needed to completely define the solution and where each irrelevant dimension would yield exactly 50% correct responses. The effects of outcomes and stimulus sequences were measured by the number of correct solutions on trial 4, given various changes on trials 1, 2, and 3 (Experiment III was of the same nature except that 16 trials were used and internal segments of two or more were studied). From experiment I, Richter concluded that the sampling with replacement axiom had to be rejected since the probability of solution on trial 4 given an error on trial 3 (P(4/ -3)) decreased as the number of errors on trials 1 and 2 increased. Further evidence indicated that the combination of an internally orthogonal set of cards along with an error on trial 3 should have resulted in a considerably lower solution rate than was obtained suggesting that subjects were not testing a single hypothesis and that sampling hypotheses

without replacement is also wrong. The final conclusion of the experiment was that when the subject is correct, information use was nearly perfect $(P(+4/(+1 \& +2 \& +3))= .975)$. Each additional error on the first three trials produced approximately equal decrements in the proportion correct on trial 4 regardless of the precise trial on which the error occurred. Experiment II was concerned with investigating subjects' efficiency in using different amounts of information and the interactions with outcomes. The results here were somewhat inconclusive although they did tend to support the findings of Experiment I. Because of this a repetition with more trials was carried out in Experiment III which allowed for a more exact estimate of solution and test of whether the learning rates for different stimulus sequences were constant across trials. The important results of this experiment indicated that subjects are influenced by rational considerations in the sense that in order to increase the probability of a correct response, they follow three redundant dimensions with a greater than chance probability. Richter concludes from Experiments I and III that not only do errors seriously impair the subject's coding or retention but that such coding or retention of information is almost perfect when the subject's choices are correct.

Levine's paper (1966) represents the first published

study concerned with correct response trials. Here using the blank trial sequence instead of Richter's method, Levine gave four feedback trials, three of which received a prearranged schedule of feedback involving the eight possible right-wrong outcome sequences. Much as in Richter's data, the results here showed that the probability of selecting the correct concept on the third blank trial set increased as the number of "right" feedbacks on the previous three feedback trials increased from 0 to 2. Levine also concludes that the subject does not just maintain his hypothesis on correct response trials but that he also stores the information of that trial and compares it with previous ones. Also subjects appear to use the outcomes to reject several incorrect hypotheses and they do this better after correct trials as opposed to error trials. Levine (1969b) later confirmed these results in a much more extensive piece of research that significantly developed his theory of information processing on correct response trials. While these latter two Levine studies may be criticized because it used the blank trial technique, the real question concerns the simultaneous presentation technique used by both Levine and Richter. In both cases each value of every dimension was present on a single trial because the subject was presented with complementary stimulus cards and asked to

name which one was correct. The cards were removed after feedback was given. Thus, the subject could predict the experimenter's feedback for the non-chosen card and as well had a short time to study this card after feedback. Wells (1967) has shown that this simultaneous contrast proceedure is easier than either the successive presentation of cards or the presentation of two cards of the same category.

Nahinsky's work, which was mentioned earlier, resembles that of Levine and Richter in many respects. He generally uses a system of a few feedback instances and non-instances with blank trials following the feedback instances (Nahinsky, 1968, 1970; Nahinsky and Slaymaker, 1969). One of his interests has centered around detecting information processing on correct response trials for conjunctive problems. He concludes that such processing does occur because the subjects making 0 or 1 errors in categorization turn out to be significantly higher in number than expected by chance. The problems in his study that weaken his arguments concern the fact that too many blank trials were classified positive, the problems were very simple (only three possible hypotheses existed on the first card) and too many subjects guessed the correct concept at the beginning of the problem implying that

there might be a saliency effect.

A number of other studies avoided these problems by using different proceedures so that their results appear somewhat more conclusive. The Merryman et al study (1968) mentioned in connection with error performance was able to show some effects of correct responses when it pinpointed the cause of decrements in performance produced by random reinforcement. The correct response trials were in some manner lowering the learning rate. This effect was later replicated (Brown and Merryman, 1970) and found to still apply when the relevant dimension did not vary during the pretraining series involving random reinforcement. The authors state that the decrements in concept identification performance with the constant relevant dimension occurred after the RR ended because the subjects recalled that the cue had received inconsistent reinforcement.

Bourne, Guy and Wadsworth (1967) also addressed themselves to this question of the relative importance of infirmed and confirmed predictions. If, as Bower and Trabasso hypothesized, a correct response trial results in the subject retaining his previous hypothesis and learning nothing from these trials then a negative reinforcement (wrong-w) should be the most important factor. The original Bus and Bus study (1956) found that achieve-

ment of a criterion in a four category sorting task was more rapid under Right-Wrong (R-W) and nothing-Wrong (n-W) reinforcement as compared to Right-nothing (R-n) reinforcement. The Bourne study showed that this was an artifact of the percentage of rights and wrongs allowable under random guessing (1/4R, 3/4W). The subjects in the R-W and n-W groups simply received more reinforcement than the R-n group and if this was taken into consideration subjects showed about equal reliance on R and W signals.

This reliance on R signals has also been established by various investigators who switched from the more common types of dependent variables to a study of response latency changes based on the Bower and Trabasso model. Erickson et al (1966) made five predictions based on the assumptions that subjects resample with replacement on errors and that they maintain their hypotheses on correct response trials. Predictions two and three which are particularly relevant to an understanding of correct response trials, were as follows: one, latency on trials preceeding the last error should be constant and two, latency on trials following the last error should be constant and equal to the latency on trials following correct responses in the presolution phase of the problem.

While response latencies on trials after errors were found to be longer than such latencies on trials after correct responses, the fact that the latencies decreased on the criterion run and that the mean latency following correct response trials declined half again as much across trials as did latencies following errors combined to disprove predictions two and three. As Erickson states, this could be due to increased confidence as a result of a focusing of attention. However, the most interesting possibility, which he mentions, is that subjects are processing less information because they have learned something from previous trials whether they are errors or not. This could also apply to the criterion where the decrease in latencies might reflect that the subject's sample of potentially correct hypotheses was being reduced perhaps as information was acquired. Levine (1969) also studied latencies in regard to correct response trials but he was interested specifically in the criterion run. The problems were of a similar nature to those used in other experiments, the only deviation in proceedure being that the subject was required to press a button when he felt he had the correct answer, give his hypothesis and then continue on with the problem. Levine found that such a point was reached by each subject in his criterion run

and that on criterion run trials before this point latencies decreased, while after it they remained constant. Using the assumption that latency is a function of the number of hypotheses to be evaluated, Levine concludes that subjects are monitoring a subset of hypotheses and that information at the end of each trial permits some to be discarded until a correct one is discovered. Two experiments (Funk, 1972; Chatfield and Janek, 1972;) supported Levine's conclusion by finding that latencies decrease across correct response trials only when subjects sample more then one hypothesis for testing. Although this is indirect evidence, the fact is it does seem consistent with an approach emphasizing the importance of correct response trials.

Response latencies have turned out not to be the only latencies supporting the importance of correct response trials. The intertrial interval (ITI), the period between feedback and onset of the next stimulus has been extensively examined by Bourne and others (Bourne, 1963; Bourne, Guy, Dodd and Justensen, 1965; Roweton and Davis, 1968; Wells, 1970) who found that generally as the length of the ITI increases problem solving in concept identification improves. In an igenious approach using this method, Bourne, Dodd, Guy and Justensen (1968) hypothesized

that if correct response trials were not used to process information then suboptimal ITI's following correct response trials should not affect performance. The finding was that there was virtually no difference between putting optimal ITI's after errors or after correct responses. Performance improved almost as much as when they were placed after every trial. Again as in previous studies, correct response trials are shown to produce detectable effects given that appropriate methodologies and dependent variables are used. Even if these effects are not as strong as those produced by errors, the evidence presented here indicates that correct response trials or confirmed predictions deserve considerably more study than they formerly commanded.

Unfortunately this about represents the sum of studies which are relevant to the understanding of correct response trials. Quite clearly they are scattered in a variety of areas, often containing methodological errors or are somewhat indirect in their investigation (use of response latencies) as the authors were concerned with other research questions. However, when they are considered in the light of the changing theoretical role assigned to errors it seems reasonable to postulate that subjects do in fact possess some ability to process information from

correct response trials and that this information does affect final concept identifical performance.

Summary

Conclusions about any area must always remain as flexible as possible and it is with this in mind that the following statements are proposed as a summary of the theoretical and empirical evidence presented here.

Blank trials which are logically empty of information must remain relegated to a secondary position in terms of their importance for information processing. This is not intended however to mean that they can be ignored for the experimental evidence has given reasonable support to the belief that subjects react consistently to them even if it is not in a strictly logical sense. Unfortunately the specific variables causing these changes are not clearly known and concept performance can either suffer or show improvement depending on the experiment. The effects of blank trials on behavior will continue to be of major importance to those experimenters using them as methodological devices for studying such things as partial reinforcement, hypothesis tracking, etc. in concept attainment problems. The effects of blank trials will probably remain difficult to determine because of the lack of logical connection between them and the various controllable task variables.

Error trials on the other hand will probably continue

to maintain their primary position as sources of information because of the ease of detecting specific behavior changes and the success in relating them to other task variables. The experimental evidence has however forced a re-evaluation of the conclusion that they alone govern the rate and type of learning which can occur. Definitions of learning, particularly those in stochastic models will undoubtedly be required to revise their rules concerning learning only on error trials or the idea that error trials set subjects back to zero in terms of eliminating all previous information.

Correct response trials are clearly the unkown factor in concept attainment problems. While placed in a minor role by most theories, they do possess a large amount of potentially useful information even if the overt behavioral changes are not evident on them. While such a role would considerably simplify theoretical models, the few studies on concept response trials suggests that the biggest problem in this area is the lack of appropriate dependent measures. No definitive statements can be made at this time but it would appear that where such measures have been used correct response trials do affect concept attainment performance often to a considerable degree. Correct response trials thus seem to be deserving of considerably more

research in order to determine their relative importance. It is the goal of this experiment therefore to investigate subjects' ability to abstract potentially useful information from such trials.

Experimental Introduction

Any study attempting to investigate the extent of the information processing shown by subjects on correct response trials must clearly recognize and take into consideration two factors. The first of these concerns the specific stimulus sequences to be used, because it is these which carry the potentially useful information which subjects are attempting to work with. The second factor involves the selection of a measure of performance which will adequately display any possible effects caused by correct response trials. It would appear reasonable to use a dependent variable closely related to the theoretical aspects of the stimulus sequence being studies rather than the more general and traditional summary statistics such as trials or errors to criteron.

The following design was proposed as an approach which would incorporate the above two factors and thus allow for a more adequate test of information processing on correct response trials. For the purpose of this experiment it has been assumed that the basic unit being stored and manipulated by the subjects during information processing is the hypothesis. A hypothesis is defined as a combination of two stimulus attributes joined together by a logical rule which produces a single consistent

response for each different stimulus. Responses themselves are assumed to be generated on the basis of the truth table assigned to the specific logical rule in conjunction with particular stimulus attributes present in the stimulus. The total number of hypotheses available for consideration will consist of all possible combinations of any two attributes without regard to order. Each of these combinations or hypotheses will classify each stimulus in a manner consistent or inconsistent with the correct response assigned by the experimenter according to the hypothesis he has designated as correct. Thus within a set of stimuli, each hypothesis can be assigned a number indicating the number of stimuli which it incorrectly classified. Barring any saliency effects, one would predict that when a stimulus was presented and no relevant information was available, subjects would sample at random from the pool of hypotheses, showing no preference towards hypotheses having high or low rates of inconsistency. It seemed appropriate therefore to test for information processing by constructing a type of pretraining series in which the experimenter could control the number of correct responses a subject emitted before he was asked to select a new hypothesis for testing. By careful selection and analysis of the stimuli in

this pretraining series, it would be possible to construct a theoretical distribution which listed all hypotheses and their respective level of inconsistency. In this regard the maximum possible range for such a distribution would be 0 to N, where an N hypothesis would have classified all stimuli in the pretraining series wrong while a zero type hypothesis would have classified all stimuli correctly. With no information processing, the distribution of subject selected hypotheses should, within sampling limits, match that of the theoretical distribution. Any deviation from this distribution would indicate an altered sampling procedure for subjects providing support for the belief that some type of information processing had occurred. By avoiding any errors in the pretraining series, it could be concluded that this information was obtained from the correct response trials.

Since this thesis represents an attempt to show that correct response trials are important sources of information, the predictions put forward here are based on the assumption that subjects can process relevant information from correct response trials, which aids in the solving of concept attainment problems.

Predictions

The selection of hypotheses for this experiment was made with a view to achieving two major objectives. The first goal centers around the major premise of this thesis, which is the belief that subjects possess appropriate strategies for processing some of the relevant information logically available on correct response trials in order to improve hypothesis testing performance. Prediction 1 states this in a precise form for testing.

Prediction 1. The probability of subjects selecting a specific hypothesis on an error trial is directly related to the number of times that hypothesis was disproven in a series of correct response trials immediately preceeding the error trials. The more times a hypothesis is disproven the less the probability of that hypothesis being selected.

A second goal was to increase the generality of any empirical support for prediction 1 by introducing two important independent variables into the experimental design. These variables, number of problems and conceptual rule difficulty, were selected because of their strong and consistent effects on traditional measures of concept learning performance.

Most previous research has established the existence of a strong learning to learn component in concept attainment with subjects generally showing improved performance over successive problems.

Prediction 2. The amount of information processing which occurs on correct response trials increases over successive problems with subjects tending to select more consistent hypotheses on each new problem.

Conceptual rule difficulty, an equally important variable, has been shown to vary with the type of logical rule used to categorize the stimuli into exemplars and non-exemplars of the concept. While there are a variety of theoretical explanations for the changes in performance which result when different rules are used, it is clear that a stable hierarchy of rules can be constructed with conjunctive rules being the easiest and conditional rules the hardest (some minor variations have been noted due to the use of different experimental procedures).

Prediction 3. As conceptual rule difficulty increases, subjects will process less of the relevant information from a series of correct response trials and hence will show a tendency to select more inconsistent hypotheses when required to resample. In addition,

number of errors and trials to criterion will increase as rule difficulty inceases.

Because of the nature of the experimental procedure used to examine the three previous predictions, a fourth interesting test arose. Trabasso and Bower, using errors and trials to criterion as measures of performance never detected any effects of possible information processing on correct response trials. In contrast, other experimenters using more novel procedures and measures seem to have found just such a type of processing. A comparison of these two approaches became possible in this experiment which not only employed a novel measure of information processing but also allowed subjects to complete each problem and thus obtained a score for errors and trials to criterion. Prediction 4 is based on the assumption that information processing on correct response trials affects overall performance (as measured by errors or trials to criterion) and that this will be observable since an attempt has been made to maximize the opportunity for such information processing.

Prediction 4. As the number of correct response trials on which information processing can occur increases, subject will show fewer errors and trials to criterion.

This prediction only partially states the interesting comparison of novel and traditional approaches to performance measurement. The most parsimonious outcomes under these circumstances would be if the two different measures were consistent in their statement about information processing, that is either it does or does not occur. There is however the possibility that the special measure will show such processing while the traditional one will not. Such an outcome could conceivably have serious theoretical implications for the use of these more common measures in concept learning research.

Method

Design

Of utmost importance for this experiment was the necessity of employing a design reasonably similar to other concept research in order that the results would show continuity and a fair degree of generalizability. For this reason, a reception paradigm was used in which the sequence of stimuli defining each concept was predetermined by _E_ rather than by _S_, as in the selection paradigm. Each _E_ defined concept problem was broken down into two parts: a special pretraining series of correct response trials, which constituted the experimental manipulation, followed by a regular type of concept problem having the same correct answer as was employed in the pretraining series. The special measure of information processing was obtained at the end of the pretraining series while the traditional measures were obtained from performance on the concept sequence forming the second part of each problem.

The experiment involved a 4 x 3 x 2 factorial design with four levels of length of correct response sequence (0, 4, 8 or 12 feedback trials in the pretraining series) three concept types (conjunctive, disjunctive and

joint denial) and two problems for each S. Each S received only one type of concept rule for all three problems (one dummy and two test problems) and the length of the correct response sequence starting each test problem was the same. For all concept types the proportion of positive and negative cards in the pretraining and test-trial sequence was .5. Every problem began with the category which was most informative considering the type of logical rule being used (e.g. conjunctive problems are best started with a positive card, while disjunctive problems are best started with a negative card).

Subjects

The Ss were 120 females from introductory psychology who received course credit for participation. All Ss were grouped into sets of 12 so that each experimental condition received one S out of each set and thus all groups were filled at an equal rate.

Task

Each S was presented at the outset of the experiment with instructions describing one of three possible logical rules (conjunctive, disjunctive or joint denial) which were to be used as a basis for categorizing stimuli into positive and negative categories as well as for constructing new

hypotheses when old ones were proven wrong. Sixty possible two valued concepts were defined within the stimulus universe based on the six binary dimensions: six, shape, color, texture, number of figures and number of borders. The Ss were also instructed in procedures which allowed them to deal with information from positive and negative cards as well as understand how to eliminate hypotheses. Each of the three experimental problems was divided into two parts, a set of pretraining stimuli followed by a test-trial series having a solution defined by one of the three logical rules. The distinction between the pretraining and test-trial stimuli was disguised from Ss by telling them that all of the following cards made up one problem and that E would stop the problem upon solution. The problems were given to Ss consecutively and were solved without any intervening rest period. For an S's performance to be included in the final data, she had to solve all three concept attainment problems within a 1 hour period. There was no limit however on the number of test-trial stimuli which Ss could see. Subjects were tested individually with the stimulus cards of all problems being presented serially and one at a time so that previous cards were removed as new ones were shown.

The pretraining series consisted of 12 stimulus cards

shown to all Ss who received varying amounts of feedback according to the specific treatment conditions. Although the specific attributes or categories for these cards varied according to the logical rule being used, the amount and rate at which hypotheses were disproven were identical (see Appendix B and Stimuli sections for more detailed descriptions). Before the feedback part of each pretraining series began, E gave S a hypothesis to start testing instead of allowing S to select her own preferred one. Thus if an S was to receive four feedback trials, she was shown the first eight stimuli for 15 sec. each and then after Trial 8 was given a hypothesis for testing on the last four trials. All Ss were told that the purpose of the experiment was to determine how hypotheses were maintained or rejected and that the hypotheses they received might or might not be correct. One of the hypotheses E gave S (problem one) was the correct one (dummy problem) and S was able to reach the criterion of twelve correct categorizations by retaining this particular hypothesis. Any Ss in any feedback condition who failed to correctly categorize all feedback cards using the concept they were given, were eliminated from the experiment (three Ss were eliminated for this reason). In the pretraining series of problems two and three Ss were told after the twelfth card that their

current hypothesis (the one <u>E</u> gave them) was wrong and that they must select another. This choice was made after the last stimulus card had been removed in order to prevent <u>S</u>s from making selections on the basis of this card. Once <u>S</u> had selected her new hypothesis, the test trial sequence began and <u>S</u> was from this point on, allowed to test her own preferred sequence of hypotheses until she reached criterion. The criterion itself consisted of making 12 correct categorizations while using the correct hypothesis. For <u>S</u>s in the zero feedback condition, the pretraining trials were shown individually for 15 sec. as before, and after the last card, <u>E</u> told <u>S</u> that she had to select a hypothesis for testing on the rest of the stimuli. In order to set <u>S</u>s in this group on an equal basis with the feedback conditions, <u>E</u> instructed them about one hypothesis which was definitely wrong (in this case the same one which <u>E</u> had given <u>S</u>s in the feedback conditions to start the pretraining series).

The test trial stimulus sequence consisted of twenty cards, which were categorized according to the same logical rule and hypothesis as had been used in the pretraining sequence. Once <u>S</u> started on the test trial series, she continued with it until she reached criterion and was never again shown the stimuli from the pretraining series.

The basic task for each S was to classify each card into a positive or negative category as indicated by her current hypothesis and then determine, once feedback had been given if she should maintain or change her hypothesis. If the hypothesis was supported by the stimulus information, she was to respond "same" indicating that she wished to maintain her hypothesis and see another stimulus card. In the case where the stimulus information showed the hypothesis to be wrong, S was required to generate another two valued hypothesis (in some cases this new hypothesis was also inconsistent with available information but nothing was said to S about this memory error). This latter hypothesis was then tested against the following stimuli and the procedure repeated.

Each S had before her, throughout the experiment, a card describing the rule being used, how this rule worked to categorize cards into a positive or negative category and a description of a strategy involving the procedure of staying with a hypothesis as long as it was consistent with the stimulus information and shifting hypotheses only when they were proven wrong.

Stimuli and Problems (see Appendix B)

Each stimulus card consisted of six attributes (one

from each of the six binary dimensions) drawn on 3 in. x 5 in. index cards. In no case did the two attributes from the same dimension ever appear together on a stimulus card. Within a specific logical rule and across feedback levels, the same stimuli were shown in the same serial position. Across logical rules, the amount and rate of hypothesis elminiation was the same even though the specific card in the same serial position might vary in category and specific attributes depending on the particular logical rule being used. Thus for example, if the hypothesis red triangle was disproven on Trials 1, 6, 7, 9, and 11 in the conjunctive twelve feedback condition, this same hypothesis was disproven on these cards in the disjunctive and joint denial twelve feedback conditions. This similarity of hypothesis elminiation was also maintained in stimuli in the test trial series. In order to achieve this, the stimuli were constructed in the following manner. The conjunctive problem was constructed first with all cards categorized into their respective positive or negative group. Next a second set of cards was drawn up such that they had all attributes completely different from those on the conjunctive cards and each card was then given a category opposite to the one it received on the conjunctive card it was paired with. Thus if a conjunctive card was positive and had the attributes two

large red triangles with dots for texture and a double border (this card would prove 45 hypotheses wrong), the inclusive disjunctive card was negative (in order that it carry as much information as the positive conjunctive card) and had the attributes one small green square with lines for texture and a single border (with these attributes the disjunctive card eliminates the same 45 hypotheses as the conjunctive card did). The joint denial problem card simply used the same attributes as the disjunctive problem except that the categories of the cards were reversed to give the same ones used on the conjunctive problem (i.e., the equivalent joint denial card would be one small green square with lines for texture and a single border and classified as positive). This meant that if the hypothesis red triangles was correct, it would put all cards in the correct category, the conjunctive card would be positive (because it had red triangles on it), the disjunctive card would be negative (because it had neither of the two attributes) and the joint denial card would be positive (because it lacked both attributes). In all cards the same 45 hypotheses were proven wrong.

Results

The data analyses presented in this section are summarized under separate subdivisions which refer to the specific hypotheses being investigated in this experiment. The primary data for testing these predications consisted of the hypothesis selection made by S at the end of the pretraining series and the number of times this particular hypothesis was proven incorrect. There were a total of sixty different hypotheses defined in the stimulus universe each of which was logically tested exactly once on any stimulus card receiving feedback. That is, each hypothesis was shown to be consistent with a specific stimulus card if it assigned the card to the correct category (either positive or negative) designated by E or inconsistent by incorrect assignment of the card. When a hypothesis was shown to be inconsistent with the categorization of a single stimulus card, it was considered to have disproven once. It was assumed for the purpose of this experiment that the hypotheses shown to be inconsistent on positive cards were equal to such inconsistencies on negative cards in terms of their effects on information processing, as long as S correctly categorized the card. The equality of positive-negative

cards assumption has not always been supported and was introduced here only to avoid presenting Ss with unusual sequence of cards in the pretraining series such as all positive or all negative sequences.

Because there were four levels of pretraining (0, 4, 8, 12 trials), three separate lists of sixty possible hypotheses were constructed with each list indicating the number of times each hypothesis was disproven with a specific level of feedback. These three lists were considered to be theoretical distributions representing the performance of a random selection procedure and they have been designated the 4F (feedback), 8F, and 12F scoring criterions. On this basis the expected mean, standard deviation and range of inconsistency scores for the 4F, 8F, and 12F hypothesis distributions were: 2.03 ($\sigma = .95$, range 0 to 4), 4.07 ($\sigma = 1.59$, range 0 to 6), and 6.10, ($\sigma = 2.38$, range 0 to 9) respectively. Using these theoretical distributions a score was assigned to the hypothesis selected by S at the end of the pretraining series indicating the number of times that particular hypothesis had been disproven. Analysis of this data was, of course, centered on determining if the treatment conditions showed the predicted differences and how Ss performance was related to chance values.

Prediction 1, the most important of the four, stated that the greater the number of times a hypothesis was disproven in a series of correct response trials, the less the probability of that hypothesis being selected on an error trial. Basically there are two aspects of this prediction which both deserve consideration. The first clearly suggests that when Ss make their hypothesis selection after the pretraining series, they tend to choose from among those hypotheses showing lower rates of inconsistency. The second aspect of this prediction is the expectation that as the opportunity for information processing increases (i.e., the number of correct response trials increases and thus the average rate of inconsistency increases per hypothesis), Ss will show improved performance.

Table 1 shows the average number of times S selected hypotheses were disproven in the various treatment conditions when compared against their respective theoretical feedback distributions (i.e., an S selected hypothesis in the four feedback condition is compared against the theoretical 4F scoring criterion). While this table summarizes the basic data, it is inappropriate in its present form for testing part of Prediction 1 because it allows only for comparisons across problems and rules within a specific feedback level. This is clearly shown

by the fact that there are three expected means in Table 1 representing the means for the different theoretical distributions. To rescore this data in a manner more appropriate for comparison across feedback levels, a single theoretical distribution was selected and all S selected hypotheses were given a score based on how inconsistent they were according to that distribution. This approach was deemed satisfactory for the following reasons.

TABLE 1

Average Number of Times S Selected Hypotheses Were Disproven When Scored According to the 4F, 8F, or 12F Distributions

Rule Type	Feedback Level	Problem 1	Problem 2	Expected Value (for both problems 1 and 2)
	4	1.4	1.3	2.03
Conjunctive	8	2.5	1.6	4.07
	12	4.5	2.9	6.10
	4	1.2	2.1	2.03
Disjunctive	8	5.1	4.0	4.07
	12	5.8	5.0	6.10
	4	2.2	1.9	2.03
Joint Denial	8	4.2	3.9	4.07
	12	6.5	3.7	6.10

Although the specific attributes used on the pretraining series varied depending on the logical rule, the relationship between the rate and type of attributes changing was the same, thus S saw stimuli changing in the same way no matter what group they were in. Furthermore this meant that when equal feedback had been given across treatment groups, all pretraining stimuli showed the same rate and type of hypothesis inconsistency. For example, the last four trials for the 4F, 8F, and 12F groups had exactly the same hypotheses eliminated at the very same points. The result of this is that the application of a single theoretical scoring distribution created no differences between pretraining stimuli except for feedback, rule type, and problems which were in fact the treatment conditions being studied. In addition the application of one scoring distribution allowed for the inclusion of the zero feedback condition as a control group. One possible problem of interpretation did exist, however, because of the fact that while the 12F criterion included all the pretraining stiumuli, the 8F and 4F scoring criterions did not. Even though each set of four pretraining stimuli contained two positive and two negative cards, it was still possible for a hypothesis to suffer some drift in inconsistency rate as stimuli were ignored in the 8F and 4F scoring criterions. For example, a specific hypothesis might be

proven wrong quite often in one set of four cards hence its inconsistency rate would drop considerably when these four cards are left out or given no feedback. However, this same hypothesis may show very little drop in inconsistency rate for the next four cards because it would be disproven very seldom in that set. Thus dropping sets of four cards due to lack of feedback does not insure that the different hypotheses show equal drops in inconsistency rate. The term drift is applied to this differential change in hypothesis inconsistency as sets of four stimuli are dropped. Because this possibility is avoided when the 12F scoring distribution is used, interpretation of data has been based solely on the outcome of this analysis. The results for the 8F and 4F data are presented in Appendix A and as will be noted they show little difference from the 12F analysis.

Table 2 shows the average number of times _S_ selected hypotheses were disproven when scored according to the 12F criterion. The expected value for scores indicated the average rate of inconsistency for the 59 hypotheses available to _S_ for selection at the end of the pretraining series. The _F_ value for the main effect of number of feedback trials was 2.74 (3, 108) $p < .05$. None of the other treatment conditions showed a significant interaction with

feedback level, although the Feedback x Problems interaction almost reached acceptable levels, $F = 2.42$ (3, 108) $p > .05$.

Examination of Table 2 and the graphic representation of this data in Figure 1 indicates why the factor of feedback levels was significant. With certain exceptions there was generally a reduction in rate of hypothesis inconsistency as the number of feedback trials increased. This reduction was quite erratic for Problem 1 performance but became clearly defined on Problem 2 as indicated by the Feedback x Problem interaction. An interesting point to note is that while inconsistency levels were generally lowest with the conjunctive rule, the absolute reduction in inconsistency level as feedback increased from zero to twelve was roughly the same for all logical rules. As a result, there was no significant Rules x Feedback interaction as one might have expected. Two of the twenty-four groups appeared to show unusual performance levels considering the experimental predictions and the performance of other groups similar to them. There is no obvious reason why the disjunctive 4 group should have shown such a low performance on Problem 1 even though this level of inconsistency is not significantly different from the expected value. This difference is not as important however as that shown by

TABLE 2

Average Number of Times S Selected Hypotheses Were Disproven When Scored According to the 12F Theoretical Distribution

Rule Type	Feedback	Problem 1	Problem 2	Expected Value
	0	4.7	5.5	6.10
	4	4.0	3.5	6.10
Conjunctive	8	4.0	2.6	6.10
	12	4.5	2.9	6.10
	Total	17.2	14.5	
	0	7.4	6.6	6.10
	4	4.0	7.3	6.10
Disjunctive	8	7.8	5.9	6.10
	12	5.8	5.0	6.10
	Total	25.0	24.8	
	0	7.5	6.0	6.10
	4	6.9	5.9	6.10
Joint Denial	8	6.4	5.8	6.10
	12	6.5	4.0	6.10
	Total	27.3	21.7	

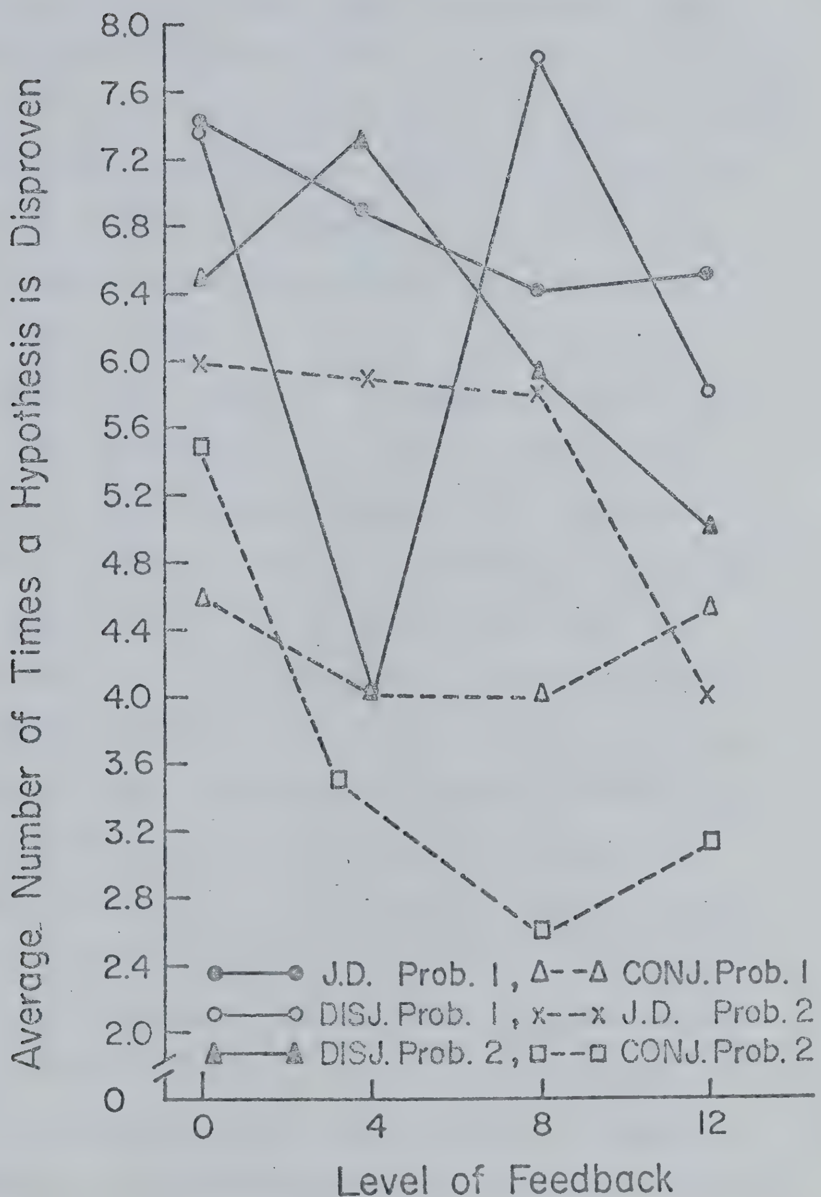
the conjunctive zero group on Problem 1. This low inconsistency rate achieved without any feedback could be caused by one of the following factors. Such a level of performance might be expected to occur occasionally as a result of sampling error, since the inconsistency rate is not significantly below the chance level. Perhaps saliency could have had certain effects which caused Ss to sample from a pool of hypotheses having a lower rate of inconsistency even though no relevant information processing occurred. On the other hand the fact that Ss are sensitive to frequency with which various attributes appear may be the cause since the relevant attributes appeared on at least six of the stimulus cards (relevant attributes would appear on the six designated positive and one half the negative cards approximately) because the conjunctive rule was being used. No choice can be made on the basis of this experiment as to which of these three alternatives is correct. Therefore because neither of these two performances were actually beyond the level expected by chance, it seems reasonable to conclude on the basis of this data that Prediction 1 is partially supported because the level of hypothesis inconsistency decreased as the number of pretraining feedback trials increased.

The above analysis was based strictly on a comparison

Figure 1

Position of Treatment Conditions When

Hypotheses Scored by the 12F Distribution



between treatment conditions. Inspection of the various tables of data indicated that group averages fell both above and below the expected value. In order to determine the effects of number of feedback trials on relative position or differences from the mean theoretical value, the data from S was scored in a special manner (these new scores to be termed the transformation data) based on the scores used in Table 1. As will be recalled the scores in Table 1 were achieved by comparing Ss with a particular level of feedback against the theoretical distribution for that feedback level (i.e., subjects receiving four feedback trials were compared against the 4F scoring criterion). Each of these scores then had the expected mean for that distribution subtracted from it and a constant value of 7.00 added to place all scores in the positive range. This set of scores represented the relative differences between treatment groups in terms of their distance from the expected mean or chance value. Since the zero feedback group had no corresponding theoretical distribution for comparison, it could either be left out of the analysis of variance or assigned a distribution and scored by it. On the basis of previous considerations, it was assumed best to use the 12F scoring criterion for the zero feedback group.

The analysis of the transformation data with the

zero feedback group left out gave an F value of 2.98 (2,80) $p > .05$ for number of feedback trials. This value was only slightly less than the 3.11 level required for significance at the .05 degree. Interestingly the Feedback x Problems F score was significant, $F = 3.69$ (2, 80) $p < .05$ and was due to the large drop in inconsistency rate in Problem 2 as feedback trials increased but very little in Problem 1. The lack of significant main effect was apparently due to the absence of zero feedback groups which in general showed the highest inconsistency rates and were the closest to the chance scores or values. With the addition of the three zero feedback groups, the F for feedback level was 6.36 (3,108) $p < .01$ indicating that it caused a significant decrease in inconsistency rate as number of feedback trials increased. The Problem x Feedback interaction was not significant here $F < 1.00$. Again, the data seems to support the conclusion that Prediction 1 is true and the lack of significance in one analysis can be explained as due to the eliminating the three most extreme groups in terms of scores.

In summary then, the analyses presented in the previous discussion support Prediction 1 in that they confirm the statement that inconsistency rate for S selected hypotheses decreases as the number of feedback

trials increases. This effect is probably modified to some extent by the other variables involved in the experiment even though the level of significance for most interactions was not quite .05. The main factor of number of feedback trials produced performance differences significant at the .05 level, therefore the actual importance of this variable in concept problems is still left open to some questioning.

In the following section consideration will be given to determining some of the characteristics of Ss' sampling which led them to show the inconsistency rates found in this experiment. Each theoretical distribution of inconsistencies has a range consisting of discrete categories which represent the number of times a hypothesis was disproven. Thus considering the 4F theoretical distribution with a range of 0 to 4, there are five categories which indicate the number of times (0, 1, 2, 3, 4) a hypothesis was inconsistent with different stimulus cards. All 59 hypotheses available to S at the end of the pretraining series fall into one of these five categories. For the 8F and 12F distributions the hypotheses fall into either 7 or 10 categories respectively. Using these categories, it is possible to compare Ss performance against the theoretical distribution to

determine at what point Ss are selecting hypotheses at lower rates of inconsistency.

Figures 2, 3 and 4 represent the percent cumulative frequency with which Ss select hypotheses from each category of inconsistency when segregated according to the theoretical distribution appropriate for scoring their particular number of feedback trials. The empirical scores from the conjunctive, disjunctive and joint denial groups were combined at each level of feedback in order to give a total of 30 scores rather than the 10 which appeared in each treatment condition. This was done in order to avoid large changes showing up when only a single score was varied.

Examination of the three graphs shows that few data points fall below the line designating the expected percent cumulative frequency obtained from the theoretical distribution. Except for part of Figure 2, Problem 2 is equal to or better than Problem 1 in the degree to which categories of lower inconsistency are selected. Furthermore it would appear that there is an increasing difference between Problem 2 and the expected distribution as the number of feedback trials goes from 4 to 8 and finally to 12. In all three figures, the empirical data seems to show a fast rise in the early categories followed by

Figure 2

Cumulative Percentage of Hypotheses Proven

Wrong for the 4F Groups

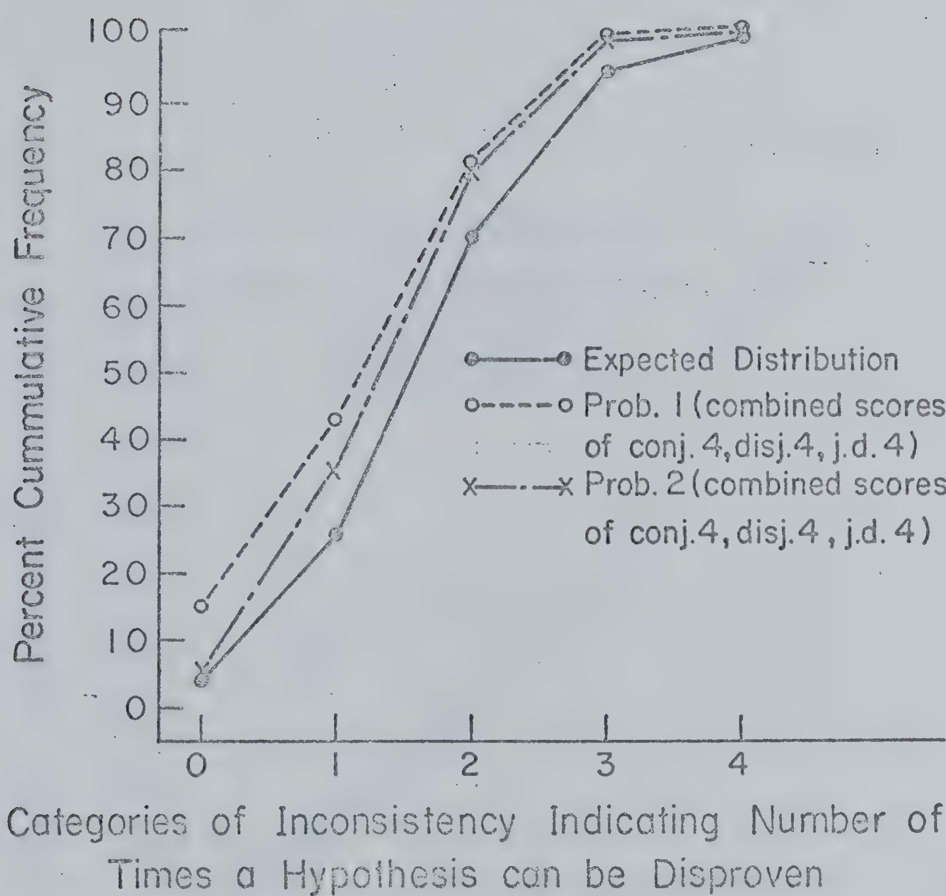


Figure 3

Cumulative Percentage of Hypotheses Proven

Wrong for the 8F Groups

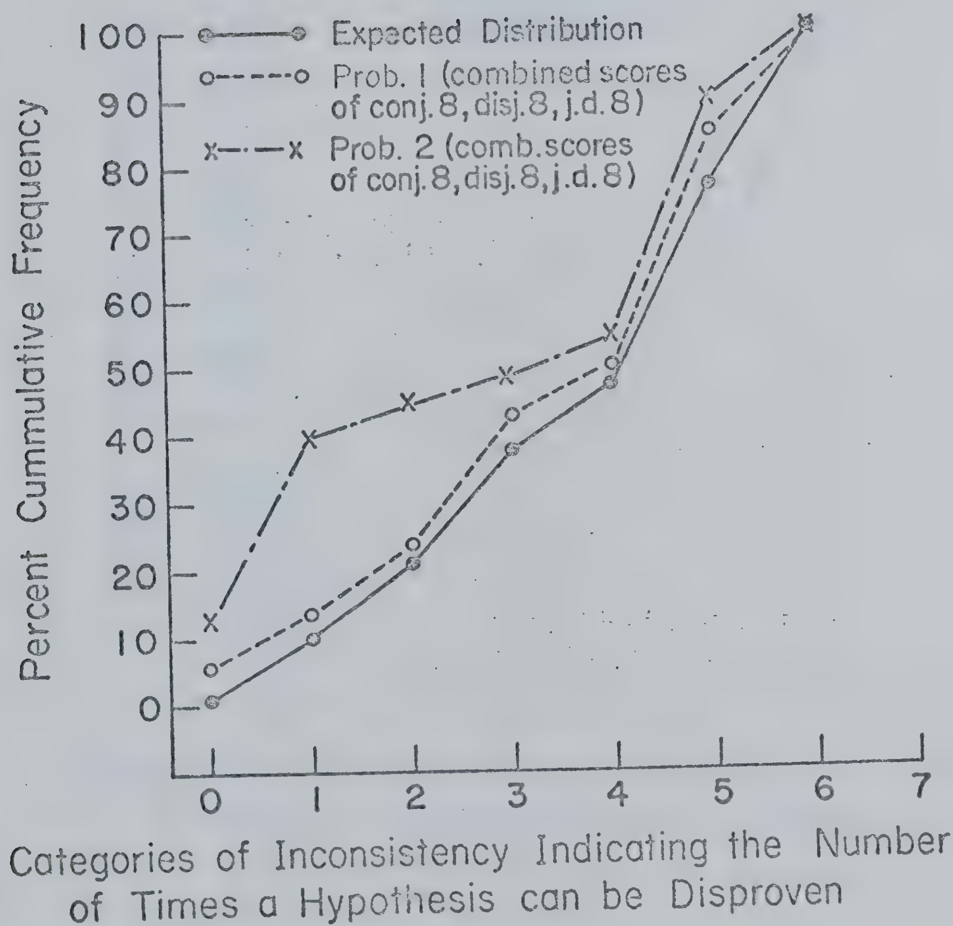
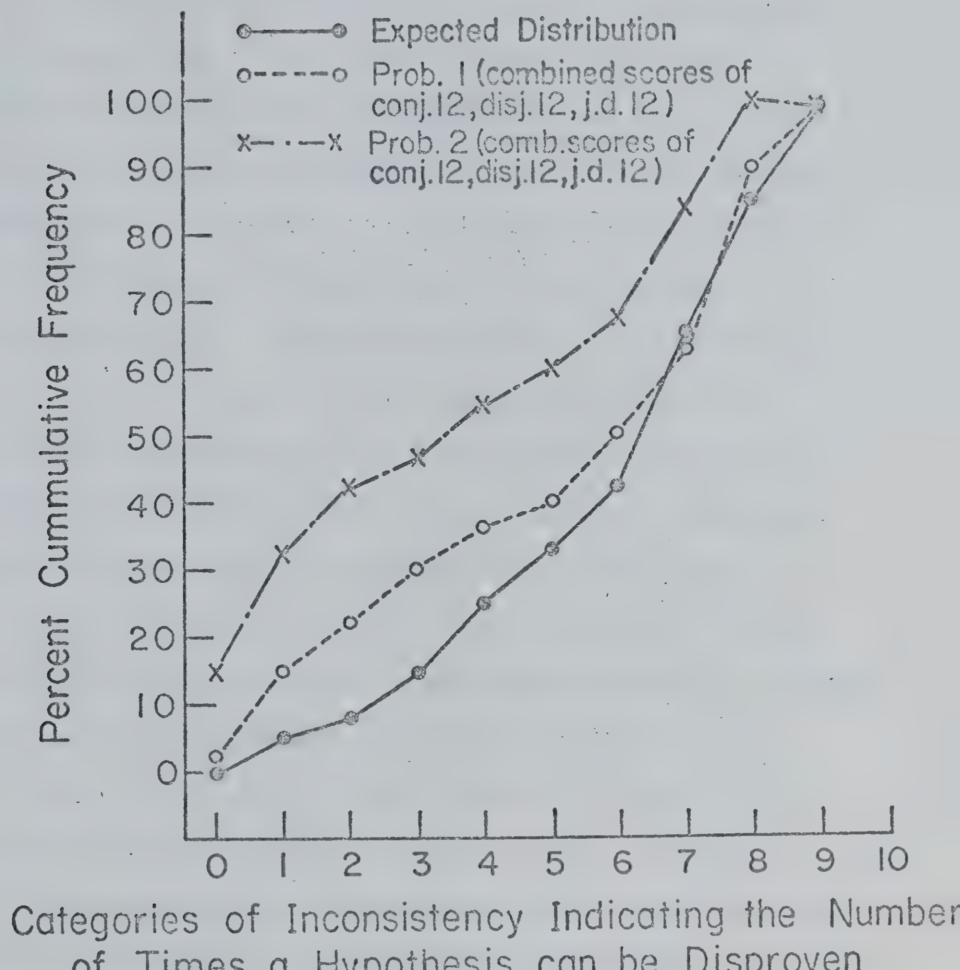


Figure 4

Cumulative Percentage of Hypotheses Proven

Wrong for the 12F Groups



a slower rate of increase until eventually the theoretical and empirical lines reach the 100% level at approximately the same point. This can be interpreted as indicating that many S selected hypotheses came from the categories with low inconsistency rates and that the remaining hypotheses are spread over the full range of categories including the ones with the highest inconsistency levels. This conclusion is confirmed by Figures 5, 6, and 7 which show the distribtuion of hypotheses in each category in terms of a percentage. With the exception of hypotheses falling in categories 0 to 1 (no inconsistencies or at most one) the distribution of Ss' selections follow that of the theoretical distribution quite closely. Subjects are not therefore showing a tendency to select from all categories of inconsistency below the average but rather only from the two lowest ones. Furthermore some Ss continue to show selections far above the mean in terms of inconsistency rate. This could be the result of faulty information processing on correct response trials. On the other hand the assumption of this experiment that all Ss can process information on correct response trials or can learn to do so may be wrong. Perhaps some Ss do learn strictly on errors and the significant effects observed here represent the performance of only a subsample of the

Figure 5

Frequency of Hypotheses Proven Wrong at
each Level of Inconsistency for the 4F Groups

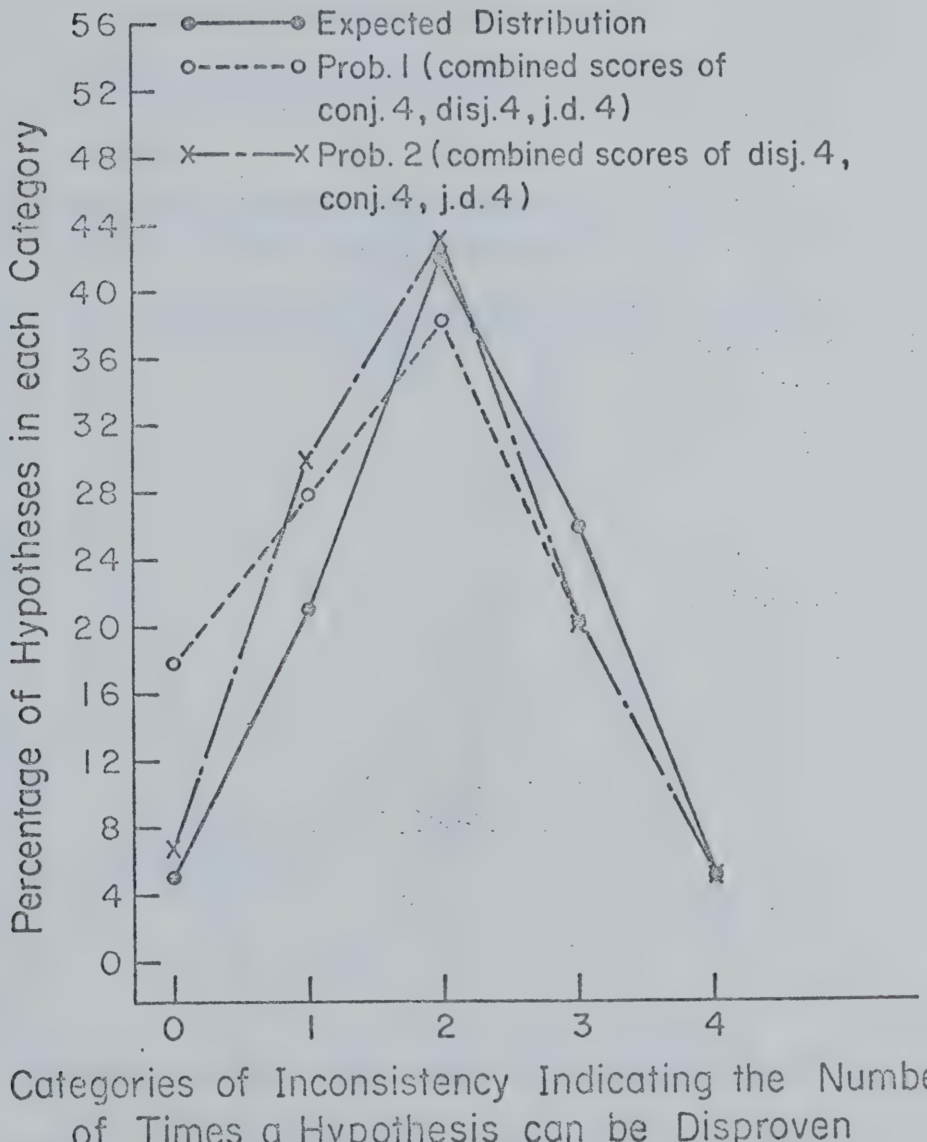


Figure 6

Frequency of Hypotheses Proven Wrong at

each Level of Inconsistency for the 8F Groups

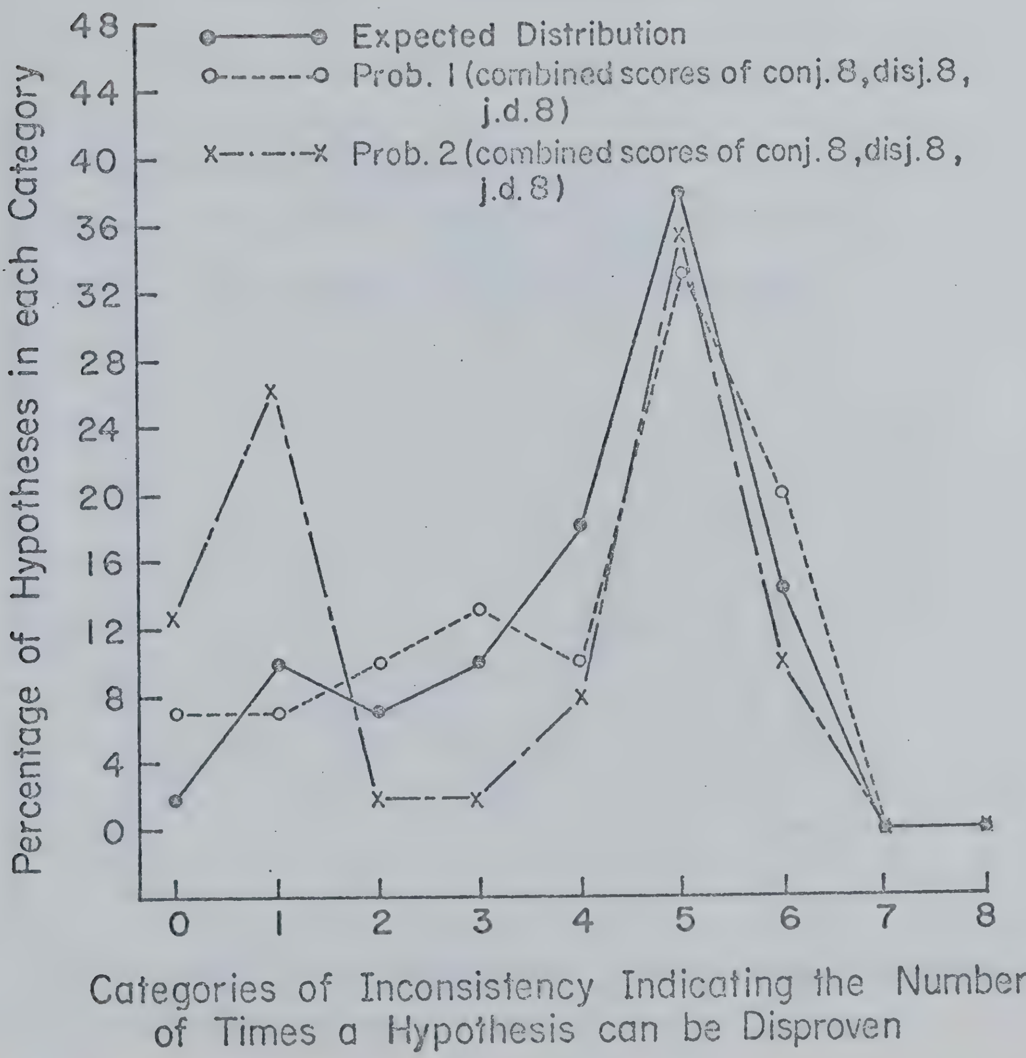
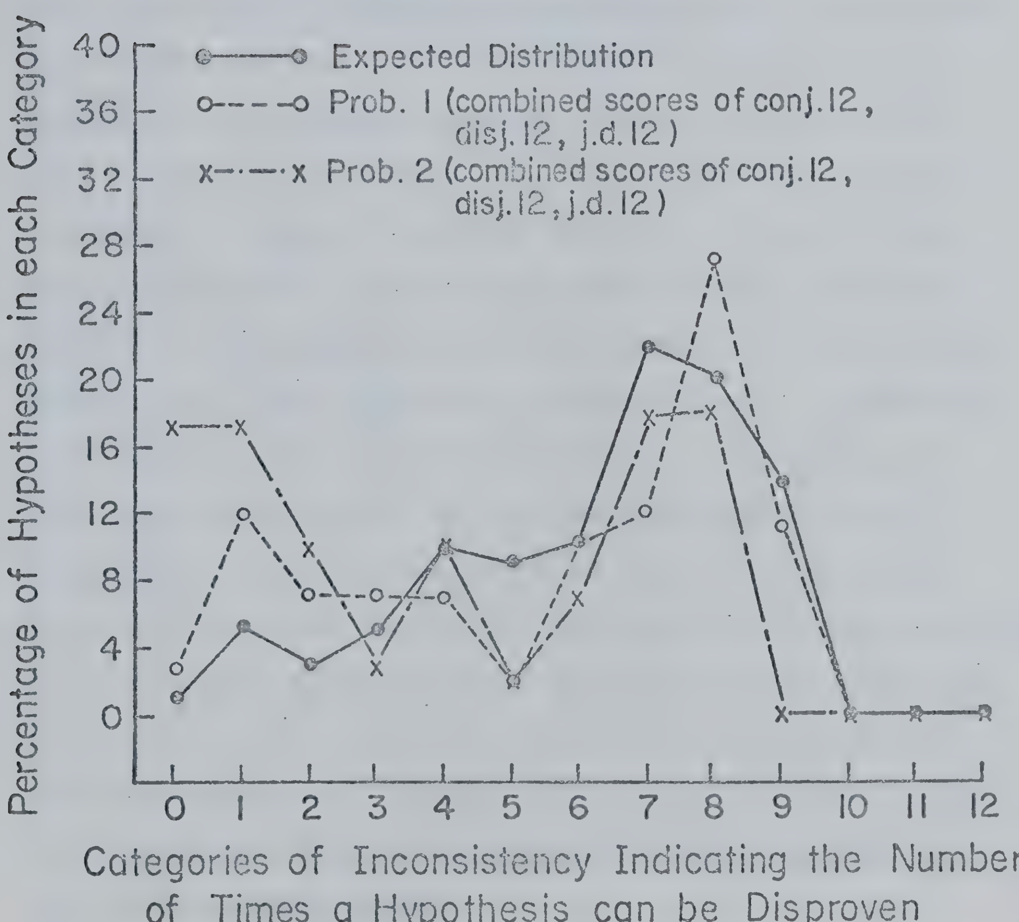


Figure 7

Frequency of Hypotheses Proven Wrong at

each Level of Inconsistency for the 12F Groups



Ss tested. These results do however lead to the conclusion that Ss tend to select hypotheses with lower rates of inconsistency than would be expected by chance, although the effect is not extremely strong.

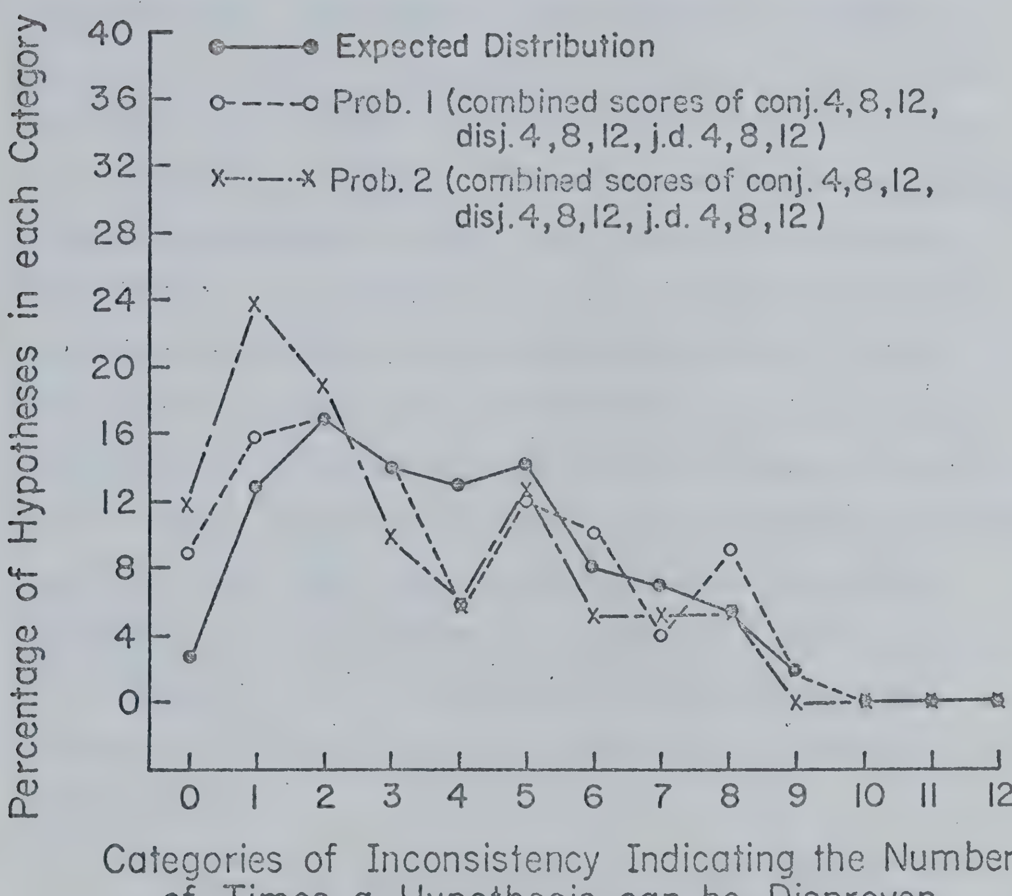
Figure 8 presents these data for all 90 Ss combined. This graph must be interpreted carefully for it represents the addition of all three theoretical distributions together.Thus the predicted number of hypotheses in the zero inconsistency category for the 4F scoring criterion are added to those in the same category for the 8F and 12F distributions. This is not done with all categories however, since certain ones do not appear in every distribution (e.g. the category of hypotheses with 9 inconsistencies appears only in the 12F theoretical distribution) thus comparisons which take into account the relative positions as some type of concrete relationship, are invalid. This simply serves to show the relationship between obtained category selections when all treatment conditions are ignored. The conclusion from Figure 8 supports the one previously drawn, namely, that Ss tend to show higher than predicted rates of selection for the categories with zero or one inconsistency.

On the basis of the results presented here it appears that Prediction 1 has received empirical support although

Figure 8

Percent Frequency Graph for all Ss Hypotheses

Irrespective of the Level of Feedback



some reservations must remain because of the failure to obtain more highly significant data levels.

Prediction 2 stated that the amount of information processing occurring on correct response trials increases over problems. Although the main effect of problems was analyzed in the data from Table 2, it must be pointed out that any analysis which includes the zero feedback groups is somewhat inappropriate. These three groups (conjunctive, disjunctive and joint denial zero) all were expected to show little or no change in their hypothesis because they received no feedback on the pretraining series. Since they did not change, the inclusion of these groups serves to diminish the significant effect of problems on the other three levels of feedback and it can be assumed that problems would have been more significant had these groups been dropped out of the analysis of variance. According to an analysis of the 12F data (Table 2) the $\underline{F}$ value for problems was 4.92 (1, 108) $\underline{p} < .05$ with no significant interactions.

Analysis of the transformation data, involving comparisons of differences from chance level, showed clearly the effects of including the zero feedback groups in the data. When these groups were excluded, problems gave an $\underline{F} = 51.7$ (1,80) $\underline{p} < .01$ with a significant

Feedback x Problems interaction $\underline{F} = 3.69$ (2, 80) $\underline{p} < .05$. However upon including them, the $\underline{F}$ value for Problems is now 3.29 (1, 108) $\underline{p} > .05$ and the Rules x Problems interaction is significant $\underline{F} = 3.34$ (2, 108) $\underline{p} < .05$. This was the result of the greater decrease across problems for conjunctive rules as compared to problems involving disjunctive or joint denial logical rules.

Figures 2, 3, and 4 give a graphic representation of the differences between Problem 2 as compared to Problem 1 hypotheses and the expected value. Generally speaking Ss make their most highly consistent hypothesis selections on Problem 2 although this improved performance is mostly limited to inconsistency levels of zero or one.

The evidence presented here seems to favor the acceptance of Prediction 2 and the conclusion that information processing on correct response trials improves over problems.

In Prediction 3 it was hypothesized that the more difficult the conceptual rule being used to categorize a set of stimuli, the less the amount of information processing which occurs and as a result, there is a greater tendency for Ss to select a wrong hypothesis after a series of correct response trials as well as commit more errors and take

more trials to reach criterion. This prediction actually contains two parts, the first to be considered here involves the consistency of hypothesis selection as affected by the rule *Ss* learned to use. The main effect of type of logical rule on hypothesis selection by *Ss* gave the strongest and most significant effect of all the experimental manipulations. Analysis of scores for the 12F scoring distribution gave an F of 15.29 (2, 108) $p < .01$. Figure 2 taken from the 12F scoring data shows the general relationship between the three logical rules. Subjects who learn conjunctive rules do significantly better in terms of selecting less inconsistent hypotheses than either the disjunctive or joint denial rules, the latter two not differing much.

Analysis of the transormation data for differences from expected mean also gave significant F scores both with and without the inclusion of the zero feedback groups $F = 9.79$ (1, 108) $p < .01$ and $F = 6.59$ (1, 80) $p < .01$. In the analysis of variance including the three zero feedback groups there was a significant Rules x Problems interaction $F = 3.34$ (2, 108) $p < .05$ which was the result of a greater decline in inconsistency on Problem 2 as compared to Problem 1 for the conjunctive rule in relationship to the other rules.

The conclusion to be drawn from this data is straight forward in regard to part A of Prediction 3. As rule difficulty increases from conjunctive to disjunctive and joint denial, S selected hypotheses are less consistent with the information available from a series of correct response feedback trials.

Part B of Prediction 3 involved an analysis of trials and errors to criterion to see if type of logical rule affected these scores. Tables 8 and 9 present the analysis of variance of all treatment groups for trials and errors to criterion respectively. The main effect of type of rule gave a significant score for both these measures: $F = 8.89$ (2, 108) $p < .01$ (trials to criterion) and $F = 9.85$ (2, 108) $p < .01$ (errors). These results are not unusual and are obtained in virtually all experiments in concept attainment which employ logical rules of different difficulty. Furthermore the similarity between the results for the two measures was expected since they are usually highly correlated, although in a few cases they do produce different levels of significance. It thus appears that part B of Prediction 3 was also supported by the experimental data.

Prediction 4 suggested that the greater the number of correct response trials on which information processing

could occur, the fewer would be the number of errors and trials to criterion when <u>S</u>s attempt to solve a concept problem. This prediction was based on the belief that, as the opportunity for information processing increased there should be a change not only in hypothesis selections but also in the more traditional measures of performance variation. This prediction was not supported by either the analysis on errors or trials to criterion, both gave <u>F</u> scores for feedback treatment which were less than one. In addition none of the interactions involving feedback was significant, the only significant variable being rules which was mentioned previously in regard to Prediction 3. It is somewhat unusual in analyses such as these to find no significant <u>F</u> for problems, because it is common to find improved performance for errors and trials over successive problems. This is probably not too important since there were only two problems used in this experiment, however it could be significant if in a replication with more problems, the improvement still was not obtained.

On the basis of this experiment, there is no evidence to support the belief that number of correct response trials affects either trials or errors to criterion in concept attainment.

In this experiment an attempt was made to determine to what degree S selected hypotheses were consistent with information available on a preceding sequence of correct response feedback trials. The support for the first three predictions, while not always of the highest significance, suggested that Ss were indeed inspecting items presented on correct response trials and processing relevant information of some type. There are however two ways in which S might have sampled hypotheses that would have possibly affected the results of this experiment. The first concerns the fact that Ss were given a hypothesis at the beginning of the pretraining series which was chosen because it was consistent with these cards even though it was eventually to be proven wrong. At the end of the pretraining series, Ss were told this hypothesis was wrong and asked to select another. It may have been that instead of processing relevant information Ss simply selected one of the two attributes from this wrong hypothesis and randomly added another. It turns out that for each problem there are eighteen hypotheses which contain one of the two attributes involved in the hypothesis originally given to Ss at the start of the pretraining series. Using the 12F scoring distribution, the approximate rate of inconsistency for

these eighteen hypotheses is 5.00, somewhat less than the expected value of 6.10. The question is therefore how many Ss actually choose such hypotheses as compared to the number expected. The eighteen hypotheses account for approximately 31% of the 59 hypotheses Ss select from. The actual percentage of Ss choosing one of these 18 was 38%, not far different from the expected value. Thus while it may have resulted in some small reduction, it does not appear that Ss were selecting their hypothesis attributes on the basis of the hypotheses given to them to start a problem to any large extent.

There is one other sampling procedure Ss might have used to improve their performance without actually processing information on correct response trials. This involves the possibility that Ss could remember the stimulus card of the pretraining series and made their hypothesis selection consistent with it. There are fourteen hypotheses consistent with the information on the last feedback stimulus of the pretraining series (excluding the one S is told is wrong but including the correct concept) having an average inconsistency rate of 4.00 approximately, which is far below the expected average of 6.10 (using the 12F scoring criterion).

Interpretation of this low rate of inconsistency is made difficult by serveral factors; however the following discussion would appear to indicate that this inconsistency rate does not in any way negate the conclusions drawn here. By the nature of the assumption regarding hypothesis sampling from the last stimulus card, it would appear that the expected consistency rate of 6.10 is an inappropriate figure for comparison. This statement is based on the following considerations. By requiring S to select a hypothesis from the last stimulus card, the experimenter is in effect changing the number of tests of consistency which that S can make for the various hypotheses. Any hypothesis appearing on the 12th or last stimulus card can only be tested for consistency on the previous 11 cards, hence the maximum number of inconsistent classifications it can show is 11. Any hypothesis not appearing on the 12th card is automatically inconsistent once and can still be tested on the previous 11 cards for a maximum number of inconsistencies of 12 not 11 as in the first case. Hence the last card is not free to vary and consistency tests must be based on the first 11 cards only. Under this condition the average rate of expected inconsistency is approximately 5.10 which

is quite a bit closer to the 4.00 level expected if Ss always select from the attributes on the last stimulus card. Furthermore if Ss were in fact making their hypothesis selections on the basis of just this last stimulus card then none of the main effects, particularly level of feedback, should cause any significant change in performance. That is all Ss with the exception of the zero feedback group should show the same performance level. Lastly the performance of several groups was far below the expected value of 4.00 thereby implying that even if this type of sampling was occurring Ss were still making hypothesis selections which required more information than was available solely on the basis of memory for the last stimulus card.

Two additional facts can be considered here although they may carry somewhat less weight. Both Trabasso (1964) and Bourne (1967) found that Ss show very poor retention capacity for stimulus attributes suggesting that the perfect memory hypothesis implied here is in fact wrong. Subjects could not process information from the last card as was considered above simply because of a very fallible memory. Secondly, even though the data is partially consistent with stimulus memory hypothesis, it is quite obvious that if Ss did process

information from correct response trials, then their selections would automatically be more consistent with the last stimulus card. In other words it is impossible to determine the direction of causality and simply because the hypotheses on the last card are less inconsistent does not imply that this was the cause of the performance obtained in the experiment.

It would seem from these arguments that there is no basis for believing that either one of these sampling procedures had a large effect on the data. Therefore the general conclusion is that Predictions 1, 2, and 3 are supported and that at least in regard to the measure used in this experiment, Ss do process valuable relevant information for testing.

Discussion

Each new study on concept learning seems to bring with it an increasingly complex and variable picture of human hypothesis testing processes. By the very nature of the results obtained here it would appear that this experiment has proven no exception to the rule. The main purpose of this study has in fact, been to show that the rather exclusive emphasis on error trails by some experimenters has resulted in over simplified theoretical explanations of a subject's ability to process potentially useful information from correct response trials. Any extensive examination of concept attainment models such as those listed by Gregg and Simon (1967) reveals that factors such as memory capacity, hypothesis sampling procedures, strategies, task demands, etc. can interact in various complex ways with both error and correct response trials. On a superficial level, performance may appear very similar between two subjects, even though they are radically different in the underlying information processing techniques they use to solve an identical problem. The answer to this problem lies in developing appropriate methodologies and measuring devices which can detect the potential differences if they exist. The experiment presented here was an

attempt to do just this, by moving beyond the overt and static behavioral nature of correct response trials, and to determine if information available in these trials affected the type of hypothesis selection subject make. While the results from this experiment supported the belief that subjects do react to correct response trials differently than most theories predict, interpretation of this data must be considered cautiously in the light of two limiting factors. First, even at this relatively primary stage of investigation into concept attainment, there have been a variety of information processing techniques postulated to explain subject's performance. The number is now well beyond the limits that can be conveniently investigated in any one study. Consequently experiments such as this one serve more to point out the subset of explanations most likely to contain the correct one rather than to state specifically which one is right. The second aspect of this problem concerns the fact that it has become increasingly easy to overgeneralize about experimental findings if sufficient attention is not given to the implicit assumptions inherent in new experimental procedures or measuring devices. Considerable attention has been paid to these two points in the following discussion in order to provide a sound

and useful basis from which to start new predictions for future research.

Since Prediction 1 formed the essential part of this base, the assumptions underlying it may be considered to be the ones most likely to place serious limitations on any conclusions that are drawn. Generally the experimental data supported Prediction 1 by finding that the inconsistency rate for subject selected hypotheses tended to decrease as the opportunity for obtaining information from additional correct response trials increased. Furthermore a division of hypotheses into subsets or categories, defined by the number of inconsistent classifications each hypothesis gave, indicated that subjects had a tendency to select from less inconsistent categories, particularly those showing no wrong classification or at the most one. Both of these results were modified to some extent by the other main effects although the reliability of these findings is low since the specific interactions involved were not quite significant at the .05 level.

Two aspects of the data suggest that a certain degree of caution must be exercised in the interpretation of this supporting evidence. First, feedback trials did not appear as a significant factor in all

the data analyses and second, the level of significance obtained by number of feedback trials was only 0.05. Since this study represents one of the few making a direct attempt to study information processing on correct response trials, the data may be considered as reasonably acceptable given the unknown nature of the processes potentially involved. Certain basic design improvements are suggested later in the discussion which could lead to a more significant display of the feedback factor. If, however, the needed replication did not result in any improvement, then a re-evaluation of the factor's importance as well as its actual existence would seem justified.

These limitations bear serious consideration in the following discussion. However, the major problem at this point concerns what conclusions about concept attainment performance may be drawn from Prediction 1, given its underlying assumptions. Quite clearly Prediction 1 requires the presence of some form of memory store which can accept the data from any information procedures which might exist. While such a memory store might appear firmly established on the basis of research in other areas of learning, the no-memory stochastic models of concept attainment had

evidence to support the belief that subjects did not possess any ability to recall stimulus sequence information or hypotheses. The lack of such capacity or even the inability of memory to accept relevant information would virtually eliminate information processing techniques as important factors in concept attainment. The evidence available in this study supports the belief that there exists relevant information that can be obtained by subjects during a period of correct response trials, which can be stored and later recalled for the purpose of modifying hypothesis selection. Because of the nature of this study, it is impossible to make any firm statements concerning certain important characteristics of memory such as the length of time information is retained or how much data can be stored at any one time. However it is possible to make some conclusions concerning the character of the information which is made available for storage. Potentially at least, the stored information could be any one or a combination of the following types: hypotheses tried and rejected on errors, stimulus attributes to be used in consistency checks, or hypotheses potentially correct and/or rejected on correct response trials. The experimental design of this study can, in conjunction with evidence from other

research, eliminate the first two possibilities from consideration.

Since no errors were allowed to occur in the pretraining series, any model which relies solely on memory for such error event information can be eliminated from consideration. Such a model cannot explain changes in subject selected hypotheses when only correct response trials are allowed to vary. Furthermore since subjects were informed of the incorrectness of their pretraining hypothesis only after all stimuli had been removed, a consistency check of the type suggested by Bower and Trabasso was not possible. The lack of an incorrectly categorized stimulus card prevented subjects from making the required comparison with a correctly classified stimulus to test for inconsistently reinforced dimensions. The only type of consistency check really possible within the context of the present experiment would be one comparing two or more correctly classified stimuli. If subjects were able to obtain information in this way, they would still be processing relevant information on correct response trials as predicted by this experiment. However, research by Trabasso (1964), and Bourne and O'Banion (1969) argues against this being likely, since subjects showed a recall level for stimulus attributues which was

only slightly above that expected by chance.

It would thus appear that probably the data available for storage in memory is in the form of hypotheses and that these are the end products of whatever information processing techniques are used by subjects. Whether these are hypotheses which subjects know are wrong or a subset of potentially right ones is unclear. The latter possibility would seem more reasonable because it is smaller and requires no transformation of data before a selection. The fact that subjects showed the greatest improvement in selection for hypotheses showing zero or one inconsistency partially supports this second possibility. If subjects do not process all information, then the above chance selection of hypotheses with one inconsistency could have occurred simply by subjects missing this single piece of data and still believing that the hypothesis was potentially correct.

Actually the evidence for storage of hypotheses only, is not quite sufficient to permanently eliminate certain other variables. Even though the theory behind Prediction 1 and the measure of information processing are based upon scoring hypotheses for their level of inconsistency it is in no sense a necessary requirement

of this experiment that subjects manipulate hypotheses. There are two variables, attribute values and stimulus dimensions, which are correlated with changes in hypotheses. With some changes in strategies, subjects could be testing the relevancy of dimensions or attributes and selecting among these while appearing to test strictly hypotheses. The wholist strategy in which a subject takes all attributes from the first plus card (conjunctive problem) and uses the other plus cards to eliminate irrelevant dimensions would be an example of such an approach. Another possibility is that there may be factors more fundamental than hypotheses which are in fact the controlling influence on hypothesis selection. Some of these were mentioned earlier in the introduction in connection with changes in subjective probabilities, however their nature is too highly speculative to deserve any further consideration at this time.

A number of theorists have disputed the former type of reasoning, for it has been their belief that testing attributes, dimensions, or any other variable for their relevance must be considered as hypothesis testing. Such an approach is not possible in this experiment since hypotheses have been defined as two

stimulus attributes joined by a logical rule. While this specific definition may lose some of the essential meaning vaguely involved in the term, it does avoid the common problem of assigning such global characteristics to the term hypothesis that it can be applied to virtually anything affecting responses. At present, hypotheses seem to have lost much of their explanatory value through over extension. In any case, the point being made is that the experimental support for Prediction 1 must not be construed as providing definitive empirical evidence for the belief that hypotheses are being processed and stored on correct response trials. The data are, however, consistent with this possibility and the hypothesis construct is considered the most logical choice because it has proven so successful in explaining other empirical data as well as being incorporated in many theories. Still, the possibility of subjects processing other variables which covary with hypotheses can not be discounted at this time. Interestingly enough much of what will be said about hypotheses can probably be applied with slight modification to these other potential variables, especially if the general definition of hypotheses is considered more appropriate.

Any consideration of memory, and the data stored there, must eventually relate back to the possible process by which that information becomes available. Because of the covert nature of these processes as well as the apparent lack of behavioral changes on correct response trials, experiments (including this one) in this area have concentrated on finding a measure of performance which would display reliable changes when certain correct response trial characteristics were manipulated. Since there is little empirical data available on which to base a description of these processes, they are commonly grouped under the title of information processing techniques. The processes that are suggested, seem usually to be based on theories or research concerned with error trials and then suitably modified to deal with the supposedly static nature of correct response trials. While the selection of these processes for investigation is not unreasonable, the empirical support for Prediction 1 would seem to indicate that they are of a more dynamic nature. It would appear that one of the major obstacles to a better understanding of these information processing techniques lies in the fact that very little is known about the actual method of response production which in the end

determines whether a trial becomes an error or correct categorization. According to most theorists, responses are produced when subjects sample a hypothesis and then use it as a basis for categorizing a stimulus. This basic assumption has been varied in a number of ways to allow for: resampling of hypotheses on errors and/or correct response trials, hypothesis sampling with limitations (e.g. with replacement, only for hypotheses consistent with the last card), mulitple hypothesis sampling with responses generated by one hypothesis, and generation of responses by samples of hypotheses. Other experimenters, while not excluding these approaches, consider the possibility that other factors may at times be the underlying cause of some responses rather than hypotheses. Subjects could, for example, emit a series of responses in order to accumulate information concerning the relevancy of dimensions (checking to see how they vary with responses and outcomes) and hence reduce the information load before testing individual hypotheses. On the other hand where correct categorization was important, subjects might take into consideration the number of dimensions changing between stimuli and maintain or switch responses depending on the degree of variation. Assuming a subject made a correct response

on trial N-1, they would make an identical response on trial N if the number of stimulus dimensions changing between trial N and N-1 was less than one-half, otherwise they would switch responses. Richter (1967) found that subjects produced responses that followed the stimulus carrying the most information more often then expected by chance. Even though there is little research presently being done on these types of responses, they do hold out the possibility of explaining some of the unusual ways in which subjects react.

This type of information gathering response, produced by variables other than hypotheses, was virtually eliminated from the pretraining series by the requirement that subjects classify all stimuli according to the hypothesis the experimenter gave them. This does not exclude such responses from occurring in the test trial sequence but since the measure of information processing occurred at the end of the pretraining sequence, it can be reasonably concluded that only such processes as affect hypothesis produced responses were measured. The theory most contradictory to the proposal of correct response trial information processing is the one that postulates a hypothetical subject who selects one hypothesis at a time for testing after an error and

generates responses from this hypothesis. On the assumption that giving a hypothesis to the subject for testing is equivalent to the subject sampling his own, then the experimental data eliminates the above possibility. During the pretraining series subjects made no errors and produced responses based on only one hypothesis and yet their performance at the end of the series indicated they were not selecting their hypotheses at random. Thus, it would appear that, single hypothesis testing, and, sampling with replacement, are eliminated from this experiment as possible processes to explain subjects' performance. In some fashion, subjects are able to take the response generated by either one hypothesis or a subset and apply it in such a manner that they can reduce the size of the hypothesis pool from which they must resample. It is interesting to note that resampling was originally assigned only to error trials because theorists felt that when a hypothesis producing the response was shown to be incorrect, subjects would be required to search for another, while on a correct respone trial subjects would not search since the hypothesis continued to work. However there is a very obvious way in which resampling might occur on correct response trials. If a subject generates responses from one hypothesis but at the same time tests

his information against a subsample then it is possible for him to eliminate and resample from this pool even though the single hypothesis is still producing correct responses. Thus if the size of the subjects subsample was six, he might after each correct response be able to eliminate on the average about three hypotheses which would have classified the stimulus wrong. These latter three would be replaced into the total pool of hypotheses while three more would be sampled to bring the subset back up to six. On any resampling the chance of including the correct concept would be low, but as the number of correct response trials (and hence the number of resamplings) increases the chance of obtaining the correct hypothesis also increases. Thus this subset would tend to accumulate hypotheses which are either correct or disproven very seldom. At the next error the subject would sample from this subset and have a better chance of picking a hypothesis which is consistent with the correct response trial information. Such techniques could give data similar to that obtained in this experiment.

In general this examination of Prediction 1 seems to lead to the following conclusions. Subjects process some relevant information on correct response trials

which is in a form that can be stored and recalled later for use in hypothesis sampling on an error. While it has not been clearly established that the processing is being carried out on hypotheses, this would appear to be the most parsimonious explanation at the present time. Quite likely subjects are able to deal with subsets of hypotheses in such a fashion that over a series of correct response trials the more inconsistent hypotheses are eliminated or receive lower subjective probabilities for sampling. The problems in the data, discussed previously, limit somewhat the confidence that can be placed on these conclusions but it does appear the subjects process information on correct response trials which is relevant to improved hypothesis selection performance.

Prediction 2 stated that the amount of information processing occurring on correct response trials increased over problems. Although only two problems were presented to each subject in this experiment, there was a significant difference in inconsistency levels for hypotheses between the two. Subject selected hypotheses on Problem 2 generally were less inconsistent with pretraining trials than were hypotheses selected on Problem 1. The same two precautions applied against Prediction 1 also

deserve to be mentioned here. Not all analyses showed a significant drop between problems and where it did occur, it was not highly significant.

Improvement in performance over problems has been a general finding in concept attainment and is considered to indicate that a learning to learn phenomena is occurring. Subjects are assumed to be focusing their attention on more relevant variables, selecting more informationally useful strategies more adequately. Support for Prediction 2 is important since it establishes that the measure of information processing used here is sensitive to a variable shown to be important in the more common types of concept attainment experiments. In addition it supports Prediction 1 by showing that the information processing techniques available on correct response trials are affected by the opportunity for learning. It seems reasonable to assume that subjects do not enter an experiment with such techniques developed to maximum capacity and thus the presentation of a series of problems allows for an improvement in handling information. One of the serious deficiencies of this experiment was the failure to include more problems in order to assess the rate of reduction in inconsistency levels. Since subjects

are not perfect information processors, the decline will eventually level out above that expected by a perfect processor, however the interesting question concerns at what point this will occur. Laughlin (1967, 1968) has found a leveling off in his experiments after about the third problem while others, working with a variety of logical rules, found small declines continuing up to about the twentieth problem. This lack of additional problems also has important implications for the interpretation of Prediction 3 as will be seen later. It would thus seem from this experiment that one can cautiously conclude that techniques for information processing on correct response trials do exist and that subjects have to learn through experience how to use them. While such learning may occur both within and between problems, this experiment could only detect the latter type.

Prediction 3 actually contains two separate parts. Part A predicts that as conceptual rule difficulty increases, subjects will process less of the relevant information from a series of correct response trials and hence will show a tendency to select more inconsistent hypotheses when required to resample. This prediction received the strongest and most consistent support throughout the experiment and thus there

can be little doubt that as the difficulty of a logical rule increases, subjects tend to select hypotheses showing more inconsistencies with pretraining sequences. This result not only serves to increase the generality of the experiment but it again serves the purpose of showing that the measure of information processing produces results much like those obtained in other concept attainment experiments which used different dependent measures. Conjunctive problems, which are generally the easiest for subjects, showed rates of hypothesis inconsistency considerably lower than those for disjunctive or joint denial problems. Furthermore, the decrease in inconsistency rates over problems was greater for the conjunctive rule than the other two. Herein lies the connection with Prediction 2, for it would have been interesting to determine if these other logical rules showed the same amount of decline after a sufficient number of problems. While some experimenters believe that the logically more complex rules will not decline as much, others have expressed the opinion that the differences are primarily the result of a lack of experience on the part of the subjects. The more difficult rules do not in any sense occupy more of the available computational space, rather,

subjects are unsure of procedures and strategies. The available evidence would seem to support the latter view since an extended series of problems usually results in little difference between rules.

Part B of Prediction 3 stated that there would be an increase in trials and errors to criterion as concept rule difficulty increased. This was supported with the finding that the conjunctive groups showed the fewest trials and errors. This result served as a check to insure that logical rules had the same effect on trials as in other experiments and that the specific procedure used in this study did not change a subject's performance in any detectable fashion.

In Prediction 4 it was hypothesizes that as the number of correct response trials on which information can occur increases, subjects will show fewer errors and trials to criterion. This prediction received no support. The number of feedback trials did not appear as a significant factor in either of these two measures. Interpretation of this finding is complicated by two results discussed previously. Prediction 1 supported the conclusion that feedback trials do affect performance at least in terms of hypothesis selection when a

special measure of information processing is used. Furthermore Prediction 3 dealing with logical rules showed that at least one well established factor affects all three measures in the same way and in a manner consistent with other experiments. This implies that there is little support for the contention that the test trial performance of subjects is unusual to any great extent. The question therefore is why the lack of effect? Many experimenters now seem to regard trials and errorsto criterion as somewhat insensitive measures of learning. The dependence of these two measures on the relative number of responses, the size of the hypothesis pool, and the length of the criterion run, etc. quite conceivably could produce variance levels which would virtually eliminate any chance of detecting differences. A second possibility however is that the variable number of feeback trials is a relatively unimportant factor and that the amount of information processing on correct response trials is quite small compared with that of error trials. Consequently the performance of subjects in the test trials series was much more affected by the errors than any information obtained from the pretraining series. The final alternative concerns the possibility that number of feedback trials

and their effect on learning cannot be measured by trials or errors because they affect different aspects of concept attainment performance. It could cause subjects to change their method of sampling hypotheses, affect their confidence in the answer or alter the degree of overlearning they show. Choosing among these alternatives to explain the lack of support for Prediction 4 is quite difficult. While the first possibility does seem to have the informal support of many experimenters, there is a lack of research data to indicate just what trials and errors do measure and the factors they are insensitive to. The third alternative is also insufficient since it is too general and theoretical. There is no obvious reason why number of feedback trials should affect completely different dependent measures at least on the basis of what is known at the present time. Therefore it has to be concluded that explanation two is the most likely possibility and that the failure to support Prediction 4 is partial evidence against the hypothesis that information processing occurs on correct response trials. At the very least it suggests that the effect is weak.

Significant consideration has been given in the previous discussion to the various theoretical problems

involved in interpreting the experimental data. At this point, however, it seems reasonable to introduce some of the specific and important limitations imposed on the results because of the particular design and measuring device used in the study.

The measure of information processing was not solely based on the effects of correct response trials. An error of sorts was required at the end of the pre-training sequence in order for subjects to abandon their current hypothesis and sample a new one. Even though this was not an error of classification (no stimulus was present), subjects still learned that their hypothesis was wrong, which is quite similar. Thus, in essence, this experiment only establishes that information obtained from correct response trials can affect how subjects re-sample on errors under certain circumstances. It does not provide any empirical data about the processing subjects might do to obtain the correct concept in a series of trials where no errors of any sort occur.

In addition since two errors never occured in a row following the pretraining series, it is impossible to determine if the information obtained on that series carries over beyond a subject's first resampling. The

failure to support Prediction 4 suggests that perhaps it does not. Conceivably Trabasso and Bower could have been partially right when they stated that error trials set subjects back to zero. Perhaps it is not the error classification itself which is important but rather the hypothesis resampling traditionally associated with errors that is the essential factor. Information obtained on the previous correct response series alters hypothesis selection but once the choice is made new data on this hypothesis and other is accumulated on the correct response trials following the error, with the old information being lost. This could be tested if subjects could in some fashion be forced to resample on a series of correct response trials at several different serial positions. If it could be shown that hypothesis selection at position 2 was poorer when a selection was required at position 1, this would perhaps indicate that information was discarded at position 1 after the hypothesis selection was made.

One final limitation on the experiment deserves to be considered even though it may be relatively minor. Because all correct response sequence occurred at the beginning of problems, there is some question about the relative effect of placing such a series at various points

throughout the problem. If subjects do obtain information from errors and/or correct response trials then the placement of a lengthy series of correct response trials later in the problem might provide little additional information. The difficulty here would be to separate out the effects of the correct response trials from the previous trials and at present the information measuring device used in this study cannot accomplish this.

These limitations suggest certain changes which could be carried out to provide additional support for the experimental predictions as well as extending the generality of the results. First more problems, possibly as many as six, should be given to each subject instead of just two. Number of problems appears to be a variable essential to the understanding of a variety of significant interactions. Such an increase in problem number would allow the experimenter to determine the degree to which subjects resemble a perfect information processor at various stages of learning. It would also permit a better test of the possibility that subjects do well on more difficult logical rules as long as they are given sufficient experience with them. Lastly it would be possible to

determine if number of feedback trials is significant on all problems or only in the early stages where subjects are showing the greatest gains in performance.

This change would permit a better understanding of inter-problem improvements. By increasing the number and position of correct response sequences within a problem there would be greater generality added to the results in this experiment. Obtaining measures on information processing under these circumstances would be difficult but it would provide basic data on the relative importance of such trials through a specific problem.

Closely related to this problem is the necessity of making some attempt to determine how long the information processed remains in memory and if it can affect hypothesis selection on errors occurring in sequences of two, three, or more. Part of this could be accomplished by simply controlling the period of time between the end of the pretraining series and the point at which subjects are told their first hypothesis is wrong. At this point the number of variations available for studying short term memory for hypotheses becomes virtually unlimited. Studies could easily be carried out to determine the effects of retroactive

inhibition either in terms of specific hypotheses or stimulus cards. Unfortunately this particular set of experiments must be set aside until the other improvements are made and the effect of correct response trials become firmly established.

In conclusion therefore it may be stated that the goal of this study was to provide conclusive evidence that subjects can process relevant information from correct response trials to be used later when making hypotheses selections. While the evidence supported this prediction, there were several important limitations which restricted the confidence that could be placed in this finding. Consequently it appears that in itself this study is not sufficient to prove that subjects process such information but that it is consistent with a growing number of experiments that have detected limited amounts of information processing on correct response trials.

The measure of information processing used in this experiment was somewhat unique but on the basis of the data would appear to be quite satisfactory. The fact that the established variables changed as predicted would indicate that the measure is reasonable and deserving of further development in regards to this and

other research questions. It now appears that this measuring device can, with suitable modifications be used to study information processing over a series of error trials thus providing a new source of information to be correlated with the vast consistent amount already obtained.

APPENDIX A

Data Analysis

TABLE 3

Analysis of Variance

Data Scored by 12F Criterion

Source	df	MS	F
Rules	2	130.74	15.29**
Feedback	3	23.46	2.74*
Rules x Feedback	6	5.21	
Error (a)	108	8.55	
Problems	1	30.11	4.92*
Rules x Problems	2	9.13	
Feedback x Problems	3	14.80	
Rules x Feedback x Problems	6	10.41	
Error (b)	108	6.12	

*$p < .05$.

**$p < .01$.

TABLE 4

Analysis of Variance

Data Scored by 8F Criterion

Source	df	MS	F
Rules	2	55.42	15.78**
Feedback	3	10.38	2.96*
Rules x Feedback	6	3.19	
Error (a)	108	3.51	
Problems	1	8.44	
Rules x Problems	2	2.84	
Feedback x Problems	3	6.49	
Rules x Feedback x Problems	6	5.92	
Error (b)	108	2.90	

*$p < .05$.

**$p < .01$.

TABLE 5

Analysis of Variance

Data Scored by 4F Criterion

Source	df	MS	F
Rules	2	10.41	10.04**
Feedback	3	2.58	
Rules x Feedback	6	1.09	
Errors (a)	108	1.03	
Problems	1	6.04	7.22**
Rules x Problems	2	.51	
Feedback x Problems	3	1.63	
Rules x Feedback x Problems	6	1.82	
Error (b)	108	.83	

$^{**}p < .01.$

TABLE 6

Analysis of Variance

Transformation Data

Zero Feedback Groups Left Out

Source	df	MS	F
Rules	2	41.95	6.59**
Feedback	2	18.99	
Rules x Feedback	4	6.91	
Error (a)	80	6.36	
Problems	1	126.30	51.70**
Rules x Problems	2	3.62	
Feedback x Problems	2	9.01	3.69*
Rules x Feedback x Problems	4	1.82	
Error (b)	80	2.44	

*$p < .05$.

**$p < .01$.

TABLE 7

Analysis of Variance

Transformation Data

Including Zero Feedback Groups

Source	df	MS	F
Rules	2	49.53	9.79**
Feedback	3	32.14	6.36**
Rules x Feedback	6	4.80	
Error (a)	108	5.06	
Problems	1	15.70	
Rules x Problems	2	15.95	3.34*
Feedback x Problems	3	.17	
Rules x Feedback x Problems	6	2.68	
Error (b)	108	4.77	

*$p < .05$.

**$p < .01$.

TABLE 8

Analysis of Variance

Number of Wrong Hypotheses

Source	df	MS	F
Rules	2	402.30	9.85**
Feedback	3	20.00	
Rules x Feedback	6	35.87	
Error (a)	108	40.84	
Problems	1	25.82	
Rules x Problems	2	44.87	
Feedback x Problems	3	14.61	
Rules x Feedback x Problems	6	21.19	
Error (b)	108	35.69	

**p < .01.

TABLE 9

Analysis of Variance

Trials to Criterion

Source	df	MS	F
Rules	2	2490.65	8.89**
Feedback	3	82.37	
Rules x Feedback	6	330.97	
Error (a)	108	280.28	
Problems	1	464.17	
Rules x Problems	2	443.81	
Feedback x Problems	3	49.92	
Rules x Feedback x Problems	6	126.60	
Error (b)	108	256.66	

**p $<$ .01.

APPENDIX B

Stimulus Construction

The stimuli for this experiment were constructed according to a number of specific rules in order to achieve certain goals (which were previously mentioned) as well as to avoid a number of theoretical and methodological problems. The specific dimensions used were chosen in order to approximate those used in other concept learning studies, while the number of dimensions (6) was selected so as to insure that the first stimulus card of each problem always had a large number of potentially correct hypotheses. The purpose of this experiment was to show that information processing on correct response trials occurred where the problems were difficult and relativey complex and not to prove the existence of such processing on trivial problems.

The conjunctive problem was the one first constructed, with the attributes for stimulus card one chosen at random. In order to minimize the role of response probabilities, the probability of a positive card appearing on any trial was set at .5 for both the correct response sequence and the test trial sequence (this was modified slightly for the correct response according to the following rules). In order that all subjects receiving feedback should start each

problem at the same information level, it was decided to commence each problem with the category of card which was most informative given the specific rule being used (positive for conjunctive and joint denial and negative for disjunctive). This meant that for subjects in the conjunctive 12F group, card one had to be positive, but it also meant that for the conjunctive 8F group card five had to be positive (since the first four trials received no feedback) and the ninth card was positive for the conjunctive 4F group (for the disjunctive groups, the first, fifth, and ninth cards had to be negative). A further restriction was that each set of four cards 1-4, 5-8, and 9-12 must contain two positive cards and two negative cards. This was necessary because (using the conjunctive rule as an example) positive cards eliminated more hypotheses than negative cards (45 as opposed to 15). Thus if a set of four stimuli contained more or less than two positive cards, it would eliminate differential numbers of hypotheses (e.g. if stimuli 1-4 had 3 positive cards and 1 negative card a total of 3 x 45 + 1 x 15 = 150 inconsistencies would occur while if stimuli 5-8 had 1 positive and 3 negative cards there would be 1 x 45 + 3 x 15 = 90 inconsistencies). All correct

response stimuli were chosen such that both the correct answer and the hypothesis given to subjects at the start of the problem would classify all pretraining stimuli correctly. The test trial stimuli were started with a card which showed that the hypothesis given to subjects was wrong.

Once the conjunctive problem was completed a second set of 32 cards (12 pretraining and 20 test trial stimuli) was made. For each stimulus in this set the attributes were exactly the opposite to those chosen for the conjunctive problem (e.g. if card one in the conjunctive problem had red and square on it, card one in this equence had green and triangle). In addition, the category of the stimuli (positive or negative) was also the reverse of that on the conjunctive problem. This arrangement meant that in the disjunctive problem, the same hypotheses were eliminated at the same rate and position as in the conjunctive problem. This interchange of attributes and categories was carried out for all 32 cards to insure that subjects had equivalent stimuli not only for the measure of information processing but for the measures of trials and errors to criterion.

The joint denial problem used the same attributes as those on the disjunctive stimuli, however, the cards had the same category classification sequence as that on the conjunctive problem. This construction again resulted in every hypothesis having the same rate and position of inconsistency as in the conjunctive and disjunctive problems.

In order to equate problem 2 with problem 1 in terms of type and rate of hypothesis elimination, the above procedure was again followed with one additional restriction. There was a reshuffling of dimension such that each dimension on the conjunctive problem was paired with a different dimension on problem 2. Thus the dimension of color in problem 1 might be paired with shape in problem 2 so that the specific changes occurring in color (e.g. red, red, red, green, red, green) would show corresponding changes in the shape (square, square, square, triangle, square, triangle or A, A, A, B, A, B, and so on for the rest of the trials on both the dimensions). Such an approach meant that for every hypothesis in problem 1 there was another hypothesis (not with the same specific attributes) in problem 2 which showed the same type and

rate of inconsistency thus making these problems directly comparable.

The stimuli recorded in the following section give an overview of all stimuli used in the experiment, the complete set is not reproduced here for reasons of length, and because the necessary information is available here. First, since problems 1 and 2 are equal in all respects (except for the specific attributes used) only problem 1 is included in the appendix. Furthermore, the table of inconsistencies which is shown for the conjunctive problem 1 is not reproduced for the disjunctive or joint denial problem 1 because these latter two problems would have the exact same tables (the one change for the disjunctive problem would be that the positive cards would be negative and vice versa). Lastly there is no listing for the 8F or 4F groups since their stimuli and inconsistency rates can be calculated from the 12F table. For example, the stimuli for the 8F conjunctive group would simply be trials 1-4 with no feedback and then trials 5-12 with the corresponding feedback. To find the inconsistencies for the 8F group, just ignore trials 1-4 and add the rest of the inconsistencies together (for the hypothesis, Red Small, the inconsistency rate for the 8F group would be 5 instead of 8 as in the 12F group).

Stimuli

Conjunctive Problem A

(12 trials of correct responses)

Correct Answer ---------------------- Large Lines

Incorrect Answer Given to S ---------- Red Two Figures

Category	Card	Dimensions					
		Color	Size	Shape	Texture	No. of Borders	No. of Figures
Plus	1	Red	Large	Square	Lines	Single	Two
Minus	2	Green	Large	Square	Dots	Double	Two
Minus	3	Red	Small	Triangle	Dots	Single	One
Plus	4	Red	Large	Triangle	Lines	Double	Two
Plus	5	Red	Large	Square	Lines	Double	Two
Minus	6	Red	Small	Triangle	Lines	Single	One
Plus	7	Red	Large	Triangle	Lines	Double	Two
Minus	8	Green	Small	Square	Lines	Double	Two
Plus	9	Red	Large	Square	Lines	Double	Two
Minus	10	Green	Large	Triangle	Dots	Single	Two
Minus	11	Red	Large	Triangle	Dots	Single	One
Plus	12	Red	Large	Triangle	Lines	Single	Two

Test Trial Stimuli

Conjunctive Problem A

Category	Card	Color	Size	Shape	Texture	No. of Borders	No. of Figures
		Dimensions					
Plus	13	Green	Large	Triangle	Lines	Double	One
Minus	14	Green	Small	Square	Dots	Double	Two
Minus	15	Green	Large	Square	Dots	Single	One
Minus	16	Green	Small	Triangle	Dots	Double	Two
Plus	17	Green	Large	Square	Lines	Double	One
Minus	18	Red	Small	Square	Lines	Single	One
Minus	19	Green	Large	Triangle	Dots	Single	One
Plus	20	Red	Large	Triangle	Lines	Single	Two
Plus	21	Green	Large	Triangle	Lines	Single	Two
Plus	22	Red	Large	Square	Lines	Single	One
Plus	23	Green	Large	Triangle	Lines	Double	One
Minus	24	Red	Small	Square	Lines	Single	Two
Minus	25	Red	Small	Square	Dots	Double	Two
Plus	26	Red	Large	Square	Lines	Double	One
Minus	27	Red	Large	Triangle	Dots	Double	Two
Plus	28	Green	Large	Square	Lines	Single	One
Plus	29	Green	Large	Square	Lines	Double	Two
Minus	30	Green	Small	Triangle	Dots	Double	One
Plus	31	Red	Large	Triangle	Lines	Double	One
Minus	32	Red	Large	Square	Dots	Single	Two

Inconsistencies for Conjunctive Problem A

Hypothesis	Positive Cards						Negative Cards						Total
	1	4	5	7	9	12	2	3	6	8	10	11	
Red Large												x	1
Red Small	x	x	x	x	x	x		x	x				8
Red Triangle	x		x		x			x	x			x	6
Red Square		x		x		x							3
Red Dots	x	x	x	x	x	x		x				x	8
Red Lines									x				1
Red Double	x					x							2
Red Single		x	x	x	x			x	x			x	7
Red 1 Figures	x	x	x	x	x	x		x	x			x	9
Red 2 Figures	Incorrect concept given to S												0
Green Large	x	x	x	x	x	x	x				x		8
Green Small	x	x	x	x	x	x				x			7
Green Triangle	x	x	x	x	x	x					x		7
Green Square	x	x	x	x	x	x	x			x			8
Green Dots	x	x	x	x	x	x	x				x		8
Green Lines	x	x	x	x	x	x				x			7
Green Double	x	x	x	x	x	x	x			x			8
Green Single	x	x	x	x	x	x					x		7
Green 1 Figures	x	x	x	x	x	x							6
Green 2 Figures	x	x	x	x	x	x	x			x	x		9
Large Triangle	x		x		x						x	x	5
Large Square		x		x		x	x						4

Inconsistencies for Conjunctive Problem A

(Continued)

Hypothesis	Positive Cards						Negative Cards						Total
	1	4	5	7	9	12	2	3	6	8	10	11	
Large Dots	x	x	x	x	x	x	x				x	x	9
Large Lines	Correct Concept												0
Large Double	x					x	x						3
Large Single		x	x	x	x						x	x	6
Large 1 Figure	x	x	x	x	x	x						x	7
Large 2 Figures							x				x		2
Small Triangle	x	x	x	x	x	x		x	x				8
Small Square	x	x	x	x	x	x				x			7
Small Dots	x	x	x	x	x	x		x					7
Small Lines	x	x	x	x	x	x			x	x			8
Small Double	x	x	x	x	x	x				x			7
Small Single	x	x	x	x	x	x		x	x				8
Small 1 Figure	x	x	x	x	x	x		x	x				8
Small 2 Figures	x	x	x	x	x	x				x			7
Triangle Dots	x	x	x	x	x	x		x			x	x	9
Triangle Lines	x		x		x				x				4
Triangle Double	x		x		x	x							4
Triangle Single	x	x	x	x	x			x	x		x	x	9
Triangle 1 Fig.	x	x	x	x	x	x		x	x			x	9
Triangle 2 Fig.	x		x		x						x		4
Square Dots	x	x	x	x	x	x	x						7

Inconsistencies for Conjunctive Problem A

(Continued)

Hypothesis	Postive Cards						Negative Cards						Total
	1	4	5	7	9	12	2	3	6	8	10	11	
Square Lines		x		x		x				x			4
Square Double	x	x		x		x	x			x			6
Square Single		x	x	x	x	x							5
Square 1 Figure	x	x	x	x	x	x							6
Square 2 Figures		x		x		x	x			x			5
Dots Double	x	x	x	x	x	x	x						7
Dots Single	x	x	x	x	x	x		x			x	x	9
Dots 1 Figure	x	x	x	x	x	x		x				x	8
Dots 2 Figures	x	x	x	x	x	x	x				x		8
Lines Double	x					x				x			3
Lines Single		x	x	x	x				x				5
Lines 1 Figure	x	x	x	x	x	x			x				7
Lines 2 Figures										x			1
Double 1 Figure	x	x	x	x	x	x							6
Double 2 Figures	x					x	x			x			4
Single 1 Figure	x	x	x	x	x	x		x	x			x	9
Single 2 Figures		x	x	x	x						x		5

Stimuli

Disjunctive Problem A

(12 trials of correct responses)

Correct Answer -------------------------- Large Lines

Incorrect Answer Given to S ------------- Red Two Figures

Category	Card	Color	Size	Shape	Texture	No. of Borders	No. of Figures
		Dimensions					
Minus	1	Green	Small	Triangle	Dots	Double	One
Plus	2	Red	Small	Triangle	Lines	Single	One
Plus	3	Green	Large	Square	Lines	Double	Two
Minus	4	Green	Small	Square	Dots	Single	One
Minus	5	Green	Small	Triangle	Dots	Single	One
Plus	6	Green	Large	Square	Dots	Double	Two
Minus	7	Green	Small	Square	Dots	Single	One
Plus	8	Red	Large	Triangle	Dots	Single	One
Minus	9	Green	Small	Triangle	Dots	Single	One
Plus	10	Red	Small	Square	Lines	Double	One
Plus	11	Green	Small	Square	Lines	Double	Two
Minus	12	Green	Small	Square	Dots	Double	One

Test Trial Stimuli

Disjunctive Problem A

Category	Card	Dimensions					
		Color	Size	Shape	Texture	No. of Borders	No. of Figures
Minus	13	Red	Small	Square	Dots	Single	Two
Plus	14	Red	Large	Triangle	Lines	Single	One
Plus	15	Red	Small	Triangle	Lines	Double	Two
Plus	16	Red	Large	Square	Lines	Single	One
Minus	17	Red	Small	Triangle	Dots	Single	Two
Plus	18	Green	Large	Triangle	Dots	Double	Two
Plus	19	Red	Small	Square	Lines	Double	Two
Minus	20	Green	Small	Square	Dots	Double	One
Minus	21	Red	Small	Square	Dots	Double	One
Minus	22	Green	Small	Triangle	Dots	Double	Two
Minus	23	Red	Small	Square	Dots	Single	Two
Plus	24	Green	Large	Triangle	Dots	Double	One
Plus	25	Green	Large	Triangle	Lines	Single	One
Minus	26	Green	Small	Triangle	Dots	Single	Two
Plus	27	Green	Small	Square	Lines	Single	One
Minus	28	Red	Small	Triangle	Dots	Double	Two
Minus	29	Red	Small	Triangle	Dots	Single	One
Plus	30	Red	Large	Square	Lines	Single	Two
Minus	31	Green	Small	Square	Dots	Single	Two
Plus	32	Green	Small	Triangle	Lines	Double	One

Stimuli

Joint Denial Problem A.

(12 trials of correct responses)

Correct Answer -------------------------- Large Lines

Incorrect Answer Given to S ------------ Red Two Figures

Category	Card	Dimensions Color	Size	Shape	Texture	No. of Borders	No. of Figures
Plus	1	Green	Small	Triangle	Dots	Double	One
Minus	2	Red	Small	Triangle	Lines	Single	One
Minus	3	Green	Large	Square	Lines	Double	Two
Plus	4	Green	Small	Square	Dots	Single	One
Plus	5	Green	Small	Triangle	Dots	Single	One
Minus	6	Green	Large	Square	Dots	Double	Two
Plus	7	Green	Small	Square	Dots	Single	One
Minus	8	Red	Large	Triangle	Dots	Single	One
Plus	9	Green	Small	Triangle	Dots	Single	One
Minus	10	Red	Small	Square	Lines	Double	One
Minus	11	Green	Small	Square	Lines	Double	Two
Plus	12	Green	Small	Square	Dots	Double	One

Test Trial Stimuli

Joint Denial Problem A

Category	Card	Dimensions					
		Color	Size	Shape	Texture	No. of Borders	No. of Figures
Plus	13	Red	Small	Square	Dots	Single	Two
Minus	14	Red	Large	Triangle	Lines	Single	One
Minus	15	Red	Small	Triangle	Lines	Double	Two
Minus	16	Red	Large	Square	Lines	Single	One
Plus	17	Red	Small	Triangle	Dots	Single	Two
Minus	18	Green	Large	Triangle	Dots	Double	Two
Minus	19	Red	Small	Square	Lines	Double	Two
Plus	20	Green	Small	Square	Dots	Double	One
Plus	21	Red	Small	Square	Dots	Double	One
Plus	22	Green	Small	Triangle	Dots	Double	Two
Plus	23	Red	Small	Square	Dots	Single	Two
Minus	24	Green	Large	Triangle	Dots	Double	One
Minus	25	Green	Large	Triangle	Lines	Single	One
Plus	26	Green	Small	Triangle	Dots	Single	Two
Minus	27	Green	Small	Square	Lines	Single	One
Plus	28	Red	Small	Triangle	Dots	Double	Two
Plus	29	Red	Small	Triangle	Dots	Single	One
Minus	30	Red	Large	Square	Lines	Single	Two
Plus	31	Green	Small	Square	Dots	Single	Two
Minus	32	Green	Small	Triangle	Lines	Double	One

References

Andrews, O.E., Levinthal, C.C., & Fishbein, H.D. The organization of hypothesis testing behaviour in concept identification tasks. American Journal of Psychology, 1969, 82, 523-530.

Bourne, L.E. Long term effects of misinformative feedback upon concept identification. Journal of Experimental Psychology, 1963, 65 (2), 139-147.

Bourne, L.E., & Bunderson, C.V. Effects of delay of informative feedback and length of postfeedback interval on concept identification. Journal of Experimental Psychology, 1963, 65 (1), 1-5.

Bourne, L.E., Dodd, D.H., Guy, D.E., & Justesen, D.R. Response-contingent intertrial intervals in concept identification. Journal of Experimental Psychology, 1968, 76 (4), 601-608.

Bourne, L.E., Guy, D.E., Dodd, D.H., Justesen, D.R. Concept identification: The effects of varying length and informational components of the intertrial interval. Journal of Experimental Psychology, 1965, 69 (6), 624-629.

Bourne, L.E., Guy, D.E., & Wadsworth, N. Verbal-reinforcement combinations and the relative frequency of informative feedback in a card sorting task. Journal Experimental Psychology, 1967, 73 (2), 220-226.

Bourne, L.E., & Haygood, R.C. Effects of intermittent reinforcement of an irrelevant dimension and task complexity upon concept identification. Journal of Experimental Psychology, 1960, 60 (6), 371-375.

Bourne, L.E., & O'Banion, K. Memory for individual events in concept identification. Psychonomic Science, 1969, 16 (2), 101-103.

Bourne, L.E., & Pendleton, R.B. Concept identification as a function of completeness and probability of information feedback. Journal of Experimental Psychology, 1958, 56, 413-420.

Bourne, L.E., & Restle, F.A. Mathematical theory of concept learning. Psychological Review, 1959, 66, 278-296.

Bower, G., & Trabasso, T. Reversals prior to solution in concept identification. Journal of Experimental Psychology, 1963, 66 (4), 409-418. (a)

Bower, G., & Trabasso, T. Concept identification. In R. C. Atkinson (ed.), Studies in Mathematical Psychology. Stanford: Stanford University Press, 1963. (b)

Brown, E.R., & Merryman, C.T. Effects of noncontingent "rights and random reinforcements on concept identification as a function of the relevant dimension's cue value". Psychonomic Science, 1970, 19 (4), 197-198.

Buss, A.H., & Buss, E.H. The effect of verbal reinforcement combinations on conceptual learning. Journal of Experimental Psychology, 1956, 52, 283-287.

Cahoon, R.L. Concept attainment and knowledge of results. Journal of Psychology, 1970, 74, 219-229.

Chatfield, D.C., & Janek, E.J. Attribute selection in concept identification. Journal of Experimental Psychology. 1972, 95, 97-101.

Chumbley, J. Hypothesis memory in concept learning. Journal of Mathematical Psychology, 1969, 6, 528-540.

Dodd, D.H., & Bourne, L.E. Tests of some assumptions of a hypothesis testing model of concept identification. Journal of Experimental Psychology, 1969, 80, 69-72.

Erickson, J.R. Hypothesis sampling in concept identification. Journal of Experimental Psychology, 1968, 76, 12-18.

Erickson, J.R., Block, K.K., & Rulon, M.J. Some characteristics of hypothesis sampling in concept identification. Psychonomic Science, 1970, 20 (2), 103-104.

Erickson, J.R., & Zajkowski, M.M. Learning several concept identification problems concurrently: A test of the sampling with replacement assumption. Journal of Experimental Psychology, 1967, 74, 212-218.

Erickson, J.R., Zajkowski, M.M., & Ehnmann, E.D. All or none-assumption in concept identification: An analysis of latency data. Journal of Experimental Psychology, 1966, 72, 690-697.

Falmagne, R. Construction of a hypothesis for concept identification. Journal of Mathematical Psychology, 1970, 7, 60-96.

Fink, R.T. Response latency as a function of hypothesis-testing strategies in concept identification. Journal of Experimental Psychology, 1972, Vol. 95, (2), 337-342.

Fishbein, H., Benton, J., Osborne, M., & Wise, H. Effects of a "Wrong" on the rejection of hypotheses and dimensions in concept learning tasks. Preceedings, 75 Annual Convention, APA, 1967, 45-46.

Gregg, L.W., & Simon, H.A. Process models and stochastic theories of simple concept formation. Journal of Mathematical Psychology, 1967, 4, 246-276.

Holstein, S.B., & Premack, D. On the different effects of random reinforcement and presolution reversal on human concept identification. Journal of Experimental Psychology, 1965, 70, 335-337.

Johannsen, W.J. Concept identification under misinformative and subsequent informative feedback conditions. Journal of Experimental Psychology, 1962, 64 (6), 631-635.

Kenoyer, C.E., & Phillips, J.L. Some direct tests of concept identification models. Psychonomic Science, 1968, 13 (4), 237-238.

Levine, M. Cue neutralization: The effects of random reinforcement upon discrimination learning. Journal of Experimental Psychology, 1962, 63, 438-443.

Levine, M. Mediating processes in humans at the outset of discrimination learning. Psychological Review, 1963, 70 (3), 254-276.

Levine, M., Leitenberg, H., & Richter, M. The blank trials law: The equivalence of positive reinforcement and non reinforcement. Psychological Review, 1964, 71 (2), 94-103.

Levine, M. Hypothesis behavior by humans during discrimination learning. Journal of Experimental Psychology, 1966, 71 (3), 331-338.

Levine, M. Latency-choice discrepancy in concept learning. Journal of Experimental Psychology, 1969, 81, 1-3.(a)

Levine, M. Neo-noncontinuity theory. In G.H. Bower & J.T. Spence (Eds.), The Psychology of Learning & Motivation. Vol. 3, New York: Academic Press 1969.(b)

Merryman, C., Kaufmann, B., Brown, E., & Dames, J. Effects of "Rights" and "Wrongs" on concept identification. Journal of Experimental Psychology, 1968, 76 (1), 116-119.

Nahinsky, I.D. A test of axioms of all or none concept identification models. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 593-601.

Nahinsky, I.D. A hypothesis sampling model for conjunctive concept identification. Journal of Mathematical Psychology, 1970, 7, 97-117.

Nahinsky, I.D., & McGlynn, F.D. Hypothesis sampling in conjunctive concept identification. Psychonomic Science, 1968, 11 (2), 77-78.

Nahinsky, I.D., Penrod, W.C., & Slaymaker, F.L. Relationship of component cues to hypotheses in conjunctive concept learning. Journal of Experimental Psychology, 1970, 83, 351-353.

Nahinsky, I.D., & Slaymaker, F.L. Sampling without replacement and information processing following correct responses in concept identification. Journal of Experimental Psychology, 1969, 81, 475-482.

Pishkin, V. Effects of probability of misinformation and number of irrelevant dimensions upon concept identification. Journal of Experimental Psychology, 1960, 59 (6), 371-378.

Restle, F.A. A theory of discrimination learning. Psychological Review, 1955, 62, 11-19.

Restle, F.A. Statistical methods for a theory of cue learning. Psychometrika, 1961, 26, 291-305.

Restle, F.A. The selection of strategies in cue learning. Psychological Review, 1962, 69, 329-343.

Richter, M. Memory, choice, and stimulus sequence in human discrimination learning. Unpublished doctoral disertation, Indiana University, 1965.

Richter, M.L. Re-evaluation of no memory results in concept identification. Journal of Experimental Psychology, 1969, 81, 591-594.

Rogers, S.P., & Haygood, R.C. Hypothesis behavior in concept identification task with probabilistic feedback. Journal of Experimental Psychology, 1968, 76, 160-165.

Roweton, W.E., & Davis, G.A. Effects of preresponse interval, postinformative feedback interval, and problem difficulty on the identification of concepts. Journal of Experimental Psychology, 1968, 78 (4). 642-645.

Trabasso, T.R. Stimulus emphasis and all-or-none learning in concept identification. Journal of Experimental Psychology, 1964, 65, 398-406.

Trabasso, T.R. Memory in concept identification. Psychonomic Science, 1964, 1, 133-134.

Trabasso, T.R., & Bower, G. Component learning in the four category concept problem. Journal of Mathematical Psychology, 1964, 1, 143-169. (a)

Trabasso, T.R., & Bower, G. Presolution reversal and dimensional shifts in concept identification. *Journal of Experimental Psychology*, 1964, 67, 398-399. (b)

Trabasso, T.R., & Bower, G. Presolution dimensional shifts in concept identification. A test of the sampling with replacement axiom in all-or-none models. *Journal of Mathematical Psychology*. 1966, 3, 163-173.

Trabasso, T.R., & Bower, G.H. Attention in learning. New York: Wiley, 1968.

Wells, H. Facilitation of concept learning by a "simultaneous contrast" procedure. *Psychonomic Science*, 1967, 9 (12), 609-610.

Wells, H. Subject-controlled intertrial intervals in concept learning. *Psychonomic Science*, 1970, 19 (2), 109-110.

Williams, G.F. A model of memory in concept learning. *Cognitive Psychology*, 1971, 2, 158-184.